

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 15:36:12 ; Search time 84 Seconds

(without alignments)
1219.112 Million cell updates/sec

Title: CACB0065

Perfect score: 2576

Sequence: 1 MSAVLVLPKAAYSVIRGSSCR.....HELLHThVAGGVVEKDEKK 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL_21:*
- 2: sp_archea:*
- 3: sp_bacteria:*
- 4: sp_fungi:*
- 5: sp_human:*
- 6: sp_invertebrate:*
- 7: sp_mammal:*
- 8: sp_mhc:*
- 9: sp_organelle:*
- 10: sp_phage:*
- 11: sp_plant:*
- 12: sp_reptile:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Query Match Length DB ID

9

09FYU1

KW	Transit peptide; Oxidoreductase.
FT	1
TRANSIT	21
CHAIN	POTENTIAL.
SEQUENCE	436 AA; 47318 MW; BF729732FB1F984 CRC64;
Query Match	54.7%; Score 1409; DB 10; Length 436; Best local Similarity 59.6%; Pred. No. 2.5e-92; Mismatches 89; Indels 39; Gaps 4; Matches 272; Conservative 56; Mismatches 89; Indels 39; Gaps 4;
QY	46 ARRLLGNVCAAAAPAAE-----APLSHVQQALAEALAKPKDDPTRKHCVCVQAPARV 97
Db	11 AGRARHVAVRAAGTSECDCPPTQAKLHWQOALDEAKPKE--SRUMIAGTASAVR 68
QY	158 HSDPLPLMPTSCCPGWIAMLEKSYPLDLYVSSCKSPONMLAANVKSLAEGKIGAPPDM 217
Db	129 NKEEPPLPMEITSCCPGWWANVERKSNEPPLIYSPQMLGAVKNNYAAQGVQPSDI 188
QY	218 VVMSIMPCTKROSEADPDMFCVADPTLRQDLYITWELGNITKERNGLAELPEGND 277
Db	189 CNVSYMPVCYRKQGADRMENTTGAGLARDVHDYTTAVGKTFLERGKLNELPESID 248
QY	278 NPMVGVSAGCAGVLFCTGGIMEALRTAVELFTGTPPLPLRSLSEVRGMDGKETNITMPA 337
Db	249 NPIGEGETGAGLIGFTGTYMEALRTVWYVTOPKPMGDFEEVRGLSCIKEATLPG 308
QY	338 PGSKFEELJKHRAAARAEEAAGHTPGPLAWDGAGFTSDRGCGITLRYAVANGIGNAK 397
Db	309 DDSFPK-----AFAGADQG-GITLKAVANGLGNAK 339
QY	398 LITNMQAGEAKYDVEIMACPACGPGGGQPRSTDQAKTOKRQALYNIKEKSPLRSHE 457
Db	340 LIKSLSEJKAQYDVEIMACPGGGQGSPRSIDQKIQKQAMYNIUDERSPIRRSID 399
QY	458 NPSTIRELYTIGLPLGHKAHELHHTYVAGGEEK 493
Db	400 NPFIQALYDKFLGAPNSHKADLHLTHYVAGGIEE 435
RESULT 4	
Q9AR66	PRELIMINARY; PRT; 449 AA.
ID	Q9AR66
AC	Q9AR66;
DT	01-JUN-2001 (TREMBlrel. 17, Created)
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE	Fe-hydrogenase precursor.
GN	HYD.
OS	Scenedesmus obliquus.
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales; Scenedesmaceae; Scenedesmus.
OX	NCBI_TAXID=3088;
RP	SEQUENCE FROM N.A.
RX	MEDLINE=2115815; PubMed=11096090;
RA	Florin L., Tsokoglou A., Happe T;
RT	"A novel type of iron hydrogenase in the green alga Scenedesmus obliquus is linked to the photosynthetic electron transport chain.";
RL	J. Biol. Chem. 276:6125-6132(2001).
DR	EMBL; AU271546; CRC34419.1; -.
DR	PS29166; IFEH.
DR	InterPro; IPR004108; Fe_hyd_19_C.
DR	InterPro; IPR003149; Fe_hyd_19_C.
DR	InterPro; IPR003149; Fe_hyd_SSU.
DR	Pfam; PF02256; Fe_hyd_SSU; 1.
DR	Protein; PF02256; Fe_hyd_19_C; 1.
DR	Pfam; PF02256; Fe_hyd_SSU; 1.
KW	Transit peptide.
FT	1
TRANSIT	35 POTENTIAL.
SEQUENCE	449 AA; 48540 MW; 9ABFC5B69580B45B CRC64;
Query Match	53.7%; Score 1384.5; DB 10; Length 449;
RESULT 5	
Q9AU60	PRELIMINARY; PRT; 403 AA.
ID	Q9AU60
AC	Q9AU60;
DT	01-JUN-2001 (TREMBlrel. 17, Created)
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE	Fe-hydrogenase (Fragment).
OS	Scenedesmus obliquus.
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales; Scenedesmaceae; Scenedesmus.
OX	NCBI_TAXID=3088;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=WILD TYPE D3;
RX	MEDLINE=21292111; PubMed=11400057;
RA	Wunschiers R./Stanger K., Senger H., Schulz R.;
RT	"Molecular evidence for a Fe-hydrogenase in the green alga Scenedesmus obliquus.";
RL	Curr. Microbiol. 42:353-360(2001).
DR	EMBL; AU2776706; AAC59621.1; -.
DR	HSSP; P28166; IFEH.
DR	InterPro; IPR004108; Fe_hyd_19_C.
DR	InterPro; IPR003149; Fe_hyd_SSU.
DR	Pfam; PF02256; Fe_hyd_19_C; 1.
DR	Pfam; PF02256; Fe_hyd_SSU; 1.
FT	NON_TER 1 1
SEQUENCE	403 AA; 43643 MW; 46110F53C137D47F CRC64;
Query Match	45.4%; Score 1196; DB 10; Length 403;
Best local Similarity 57.7%; Pred. No. 3.3e-77; Mismatches 96; Indels 38; Gaps 9; Matches 248; Conservative 48; Mismatches 96; Indels 38; Gaps 9;	
SQ	

RESULT 8

Q9XC55

ID 09XC55 PRELIMINARY; PRT; 579 AA.

AC Q9XC55; Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DR Hydrogenase-1 (SC 1.18.99.1) (Fragment).

GN HYDA.

OS Clostridium thermocellum.

OC Bacteria; Firmicutes; Bacilli; Clostridiales; Clostridiaceae; Clostridium.

OX NCBI_TAXID-1515;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 27405;

RA Desai S.G., Stevenson D.M., Prince H.L., Guerinot M.L., Lynd L.H., Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

RT "Direct Submission."

DR EMBL: AF149212; AAC35071.1; HSSP: P00195; ICP: IPR001450; 4Fe4S ferredoxin.

DR InterPro: IPR001041; Ferredoxin.

DR InterPro: IPR004108; Fe_hyd_19_C.

DR InterPro: IPR003149; Fe_hyd_SSU.

DR Pfam: PF00111; fer2; 1.

DR Pfam: PF00037; fer4; 2.

DR Pfam: PF02906; Fe_hyd_19_C; 1.

DR Pfam: PF02256; Fe_hyd_SSU; 1.

DR PRINTS: PRO0353; 4Fe4S_FERREDOXIN.

DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; 1.

KN Iron-sulfur; Oxidoreductase.

FT NON_TER 1

SEQUENCE 579 AA; 62033 MW; AEF8134FBB63B18 CRC64;

Query Match 34.7%; Score 893; DB 3; Length 555; Best Local Similarity 44.5%; Pred. No. 2.1e-55; Mismatches 181; Conservative 67; Nmatches 105; Indels 54; Gaps 7;

FT SEQUENCE 579 AA; 623584 MW; 4D5A5BD4FA526DC7 CRC64;

Query Match 35.1%; Score 903.5; DB 2; Length 579; Best Local Similarity 41.1%; Pred. No. 4e-56; Mismatches 198; Conservative 61; Nmatches 146; Indels 77; Gaps 7;

Db 251 TGRKMAALRRMGFDAYFDTNFAADLTMEGSELLERIT-----KRGK-LPMTSCSP 302

Qy 172 GWIAMLEKSYDPLIPYVSSCKSPOMMLAMVSKYLAERKGIAPKDMVMSIMPCTRKSE 231

Db 303 GWAEECKYYPERFDINSTCSPHMGKSYAEKGKIDPDKFVWSIMPCTAKKE 362

Qy 232 ADDRWFECVADPTLRLDHDVITVLELGNIFKERGINLAELPGEWDMPMGVSGAGVLF 291

Db 363 IEREMIRNG--MKDVAUTTRELARMKEMGIDFWNLKDEEFDEPLGMSTGAGAIFG 419

Qy 292 TTGGVMEALRTAYELTGTPIRPLSIEVPGSKFEELLKIRAA 351

Db 420 ATGGVMEALRTVAETVGEGRDGIKIDEFEVRGLEGYVREATTI----- 462

Db 463 -----DGMDKIALANGTGNAKKLDKYKAGEYHF 494

Qy 412 VEIMACPAGCVGGQQP--RSTDKAITQKROALYNIDEKSTLRSRSHENSPIRELYD 468

Db 495 IEVMGCGCIMGGQQTINPNMEMEEVKLRAKAEIDKNLPIRKSHENPAKRYEEF 554

Qy 469 LGEPGLGHKAHELHTY 485

Db 555 IGPPLSEKSHELLHTY 571

RESULT 9

Q8TG63

ID Q8TG63 PRELIMINARY; PRT; 555 AA.

AC Q8TG63; Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE [Fe]-hydrogenase (Fragment).

OS Piromyces sp. E2.

OC Eukaryota; Fungi; Chytridiomycota; Neocalymasticales; Neocalymastaceae; Piromyces.

OX NCBI_TAXID-73868;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-E2;

RA Vocken, F.G.J., Boxma, B., van Hoek, A.H.A.M., Athmanova, A.S., Voetels, G.D., Huynen, M., Veenhuis, M., Hackstein, J.H.P.;

RT "A hydrogenosomal [Fe]-hydrogenase from the anaerobic chytrid Neocalymastix sp. L2." Gene 284:103-112 (2002).

RL EMBL: AF446076; AAL9059.1; -.

FT NON_TER 1

SEQUENCE 555 AA; 62033 MW; AEF8134FBB63B18 CRC64;

Query Match 34.7%; Score 893; DB 3; Length 555; Best Local Similarity 44.5%; Pred. No. 2.1e-55; Mismatches 181; Conservative 67; Nmatches 105; Indels 54; Gaps 7;

FT SEQUENCE 579 AA; 623584 MW; 4D5A5BD4FA526DC7 CRC64;

Query Match 35.1%; Score 903.5; DB 2; Length 579; Best Local Similarity 41.1%; Pred. No. 4e-56; Mismatches 198; Conservative 61; Nmatches 146; Indels 77; Gaps 7;

Db 81 DPTRKHVQVAPAVRVAETGLGLPAGTPKQLAEGIRRLGFEDFDTLFGADLTME 140

Db 186 DTKKIVVYSTAPKIRVALAEEFAFPDFDTKMGKIKGFDFYIDTNTNSADLTME 245

Qy 141 EGSLILHRTEHLAHPHSDEPLMFSTSCPGWAMLEKSYDPLIPYVSSCKSPOMMLA 200

Db 246 EGTLITLRLNE-----GGKPPMFTSCPGWAMVNEKSYPIERDNLSSCSFQOMIGA 297

Qy 201 MYKSYLAERKGIAPKDMVMSIMPCTRKSEADPWFVADPTLRLDHDVITVLELGN 260

Db 298 VIKYFAKKINAKEDDIIVSVMCTAKGEAKPEKRDGV--DIDHVIITRELITL 354

Qy 69 QQALAEALKPKDPPTRKHVQVAPAVRVAETGLGLPAGTPKQLAEGLRLRGDFDFV 128

RESULT 14

Q96TP2	PRELIMINARY;	PRT;	589 AA.
AC			
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Putative 64kDa iron Hydrogenase (Fragment).		
OS	Trichomonas vaginalis		
OC	Eukaryota; Parabasidae; Trichomonidae; Trichomonas		
OX	NCBI_TaxID=572;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20523972; PubMed=11070057;		
RA	Horner D.S., Foster P.G., Embley M.T.;		
RT	"Iron Hydrogenases and the Evolution of Anaerobic Eukaryotes. ";		
RL	Mol. Biol. Evol. 17:1695-1709 (2000).		
DR	EMBL; AF262401; AAC31037.1; -.		
DR	HSPB; P29156; IFEH.		
DR	InterPro; IPR001450; 4Fe4S_ferrodoxin.		
DR	InterPro; IPR001041; ferredoxin.		
DR	InterPro; IPR004108; Fe_hyd_19_C.		
DR	InterPro; IPR003449; Fe_hyd_SSU.		
DR	Pfam; PF00111; fer4; 1.		
DR	Pfam; PF02906; Fe_hyd_19_C; 1.		
DR	Pfam; PF02256; Fe_hyd_SSU; 1.		
DR	PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.		
FT	NON_TER 1		
SEQUENCE	589 AA;	64854 MW;	FB069ED78B9D066A CRC64;

Query Match 31.3%; Score 805.5; DB 5; Length 589; Best Local Similarity 36.2%; Pred. No. 4e-49; Matches 177; Conservative 79; Mismatches 132; Indels 101; Gaps 8; QY 6 LKPCAAVSSRGSSRARQAVAPRPLAASPTVVALTEPARRLGNVACAAAPAAEPL 65 187 IQPFFGVVTIQTSC-----IKCGOCLTICPV-----GATEK 218

Db 65 SHYQOALALIAKPKDPDKTPRKHVQCVQAVAPRVAIAETGLPAGATTPKQLAEGSLRRGFD 125 219 SOVKEALDIA---NKGKRTIVQVAPAVRVALSEAFQYKEGIVTGTGRMSALKALGF 274

Db 125 EVFTLFGADLTIMEEGSSLHLRTELEAHPHSDEPLEMFSTCCPGIAMILKSYPLI 185 275 LIVDTNTYGAADLTICCEAGELVNRDR-----PNAKEPMFTCCPAPWNYVQOSAPDFI 327

Db 186 PYVSSCKSQQMLAMVSYLAERKGIAKPMVAVTMCPTKQSEADRDFWCVDADDTL 245 328 PNLSSCRSPQGMLSLAKIKNYLPLKLDVQEDVLFNSIMCTAKKDEVER-----PEL 379

QY 245 R-----QDHVTTVEGLGNFKERGINLAELGEWGNPMGCGSAGVLFGTGGYNEA 299 380 RTKSGLKEDTMVLTRELEVENMILSNIDFNNLPDQFDNIFGGSGAGQIFAATGGVME 439

QY 300 ALRPAVLELTGTPPLRSLSEVRMDGKETNITMVPAQGSKFELLKRAARAAAH 359 440 ASRPAFEVVTYKKTNTVNYPVRMDGLRLIAEDL-----PEL 474

QY 360 GTPGPLAWDDGAGFTSEDRGGTTLRVAVANGLGNNAKKLITKMQAGEAK--YDFVEIMA 416

RESULT 15

Q46606	PRELIMINARY;	PRT;	606 AA.
AC			
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	Hyd gamma.		
GN	HYDC		
OS	desulfovibrio vulgaris (strain Hildenborough).		
OC	bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae; Desulfovibrio.		
OX	NCBI_TaxID=882;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=HILDENBOROUGH;		
RA	Stokkermans J.P.W.G.;		
RL	Submitted (JAN-1991) to the EMBL/GenBank/DDBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=HILDENBOROUGH;		
RA	Stokkermans J., van Dongen W., Kaan A., Van Den Berg W., Veeder C.;		
RT	"Hyd Gamma, a gene from <i>Desulfovibrio vulgaris</i> (Hildenborough) encodes a polypeptide homologous to the periplasmic hydrogenase.";		
RL	FEMS Microbiol. Lett. 58:217-222(1989).		
DR	EMBL; X57838; CAM40970.1; -.		
DR	HSFP; P29166; IFEH.		
DR	InterPro; IPR01450; 4Fe4S_ferrodoxin.		
DR	InterPro; IPR01041; Ferredoxin.		
DR	InterPro; IPR04108; Fe_hyd_19_C.		
DR	InterPro; IPR03449; Fe_hyd_SSU.		
DR	Pfam; PF00111; fer4; 1.		
DR	Pfam; PF00037; fer4; 2.		
DR	Pfam; PF02906; Fe_hyd_19_C; 1.		
DR	Pfam; PF02256; Fe_hyd_SSU; 1.		
DR	PRINTS; PRO0353; 4FE4S_FERREDOXIN.		
DR	PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.		
DR	Iron-sulfur; 4Fe4S_FERREDOXIN.		
SQ	SEQUENCE 606 AA;	66030 MW;	F3EB14EA04C347D6 CRC64;

Query Match 30.8%; Score 794.5; DB 2; Length 606; Best Local Similarity 42.4%; Pred. No. 2.5e-48; Matches 169; Conservative 61; Mismatches 112; Indels 57; Gaps 6; QY 90 QVAPAVRVAIAETGLPAGATTPKQLAEGSLRRGFDYFDTLFGADLTIMEEGSSLHLR 149 233 QFAPAVRIGEERGLPQGSSVKGQVPAFLRAGDWDVLTNFNAADLYMEBTELQLR 293 150 TELEAHPHSDEPLEMFSTCCPGIAMILKSYPLI-----PNAKEPMFTCCPAPWNYVQOSAPDFI 209 210 KGTAKDMYMSIMPCTKQSEADRDFWCVDADPTLRLDHYTTVEGLGNFKERGINLA 269 345 MNVAPFERMVSLMPCTKARKEAPEFRRDG--VRDVAFLTREFARLKRREGIDLA 401

QY 270 BLPRGEWMDNP-MGUGSGAGVLFGTGGYMEALRTAYELFTGIPPLPRSLSEVRMDGK 328 402 GLEPSCDPLMGRATGAAVFTGGYMEALRTVYHVLNGKELAPVHLALGYENR 461

QY 329 ETNITMVPAQSKFEEELKRAARAAHAAHGTPGPLAWDDGAGFTSEDRGGTTLRVA 386

Db 462 EAVV-----
 Qy 389 ANGLIGNAKKLITKMQAGEAKYDFVEIMACPAGCVGGGQPRS - TDKAITOKRQRAILYN 446
 Db 479 VHKGIAKROVAVLAKGRADHVEMACPGGCMDDGQPRSKRANYPNQARRAALFSL 538
 Qy 447 DEKSTLRSHENISPRELYDTLGEPGLKKAHELTHY 485
 Db 539 DAENALRQSHNNPLIGKVVESLGEPCSNMSHRLHTY 577

Search completed: June 3, 2003, 16:09:07
Job time : 88 secs

This Page Blank (uspto)

Gencore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 3, 2003, 15:38:18 ; Search time 27 Seconds

Perfect score: 2576 Sequence: 1 MSALVILPKCAAVSIRGSSCR..... HELLPHHYVAGGVVEERDEKK 497

Scoring table: BLOSUM62 gapop 10.0 , Gapext 0.5

Title: CAC80065
Title: CAC80065
Run on: June 3, 2003, 15:38:18 ; (without alignments)
Scoring table: BLOSUM62
gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 20000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

Issued_Patents_AA:*

- 1: /cgp12_6/ptodata/1/1aa/5A_COMB.pep:*
- 2: /cgp12_6/ptodata/1/1aa/5B_COMB.pep:*
- 3: /cgp12_6/ptodata/1/1aa/6A_COMB.pep:*
- 4: /cgp12_6/ptodata/1/1aa/6B_COMB.pep:*
- 5: /cgp12_6/ptodata/1/1aa/PCRTUS_COMB.pep:*
- 6: /cgp12_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Query Length DB ID

Description

Result No.	Score	Query	Length	DB	ID	Description
1	125.5	4.9	748	4	US-09-056-046-4	Sequence 4, Appli
2	112	4.3	472	2	US-08-804-227C-2	Sequence 2, Appli
3	109.5	4.3	857	4	US-08-887-534A-47	Sequence 47, Appli
4	105.5	4.1	3491	2	US-07-642-734C-2	Sequence 2, Appli
5	105.5	4.1	3491	3	US-08-439-009A-2	Sequence 2, Appli
6	100.5	3.9	3170	4	US-09-036-987A-4	Sequence 4, Appli
7	100.5	3.9	3170	4	US-09-370-700-4	Sequence 4, Appli
8	99.5	3.9	638	2	US-08-125-4	Sequence 4, Appli
9	99.5	3.9	638	2	US-08-455-355-4	Sequence 4, Appli
10	99.5	3.9	3724	2	US-08-804-227C-10	Sequence 10, Appli
11	99.5	3.9	3724	2	US-08-804-198-4	Sequence 4, Appli
12	99.5	3.9	4550	2	US-08-804-227C-8	Sequence 8, Appli
13	99.5	3.9	4550	2	US-08-804-198-2	Sequence 4, Appli
14	98.5	3.8	823	4	US-09-134-001C-4081	Sequence 4081, Appli
15	94.5	3.7	1996	2	US-08-804-227C-9	Sequence 9, Appli
16	94.5	3.7	1996	2	US-08-804-198-3	Sequence 3, Appli
17	94	3.6	537	3	US-09-105-537-4	Sequence 2, Appli
18	94	3.6	805	4	US-09-513-783A-178	Sequence 178, Appli
19	94	3.6	2035	1	US-08-046-885-5	Sequence 5, Appli
20	94	3.6	2035	1	US-08-103-5	Sequence 2, Appli
21	94	3.6	2035	5	PCT-US93-11721-5	Sequence 5, Appli
22	94	3.6	4545	2	US-08-804-227C-14	Sequence 14, Appli
23	93	3.6	3782	4	US-09-105-537-4	Sequence 4, Appli
24	93	3.6	4551	3	US-09-320-878-1	Sequence 1, Appli
25	93	3.6	4613	4	US-09-105-537-31	Sequence 31, Appli
26	93	3.6	11877	4	US-09-105-537-6	Sequence 6, Appli
27	92.5	3.6	514	4	US-09-385-028-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-066-046-4
; Sequence 4, Application US/09066046A
; Patent No. 6204252
GENERAL INFORMATION:
APPLICANT: MURPHY, Cheryl
BELITZ, Gerald A.
COUGHLIN, Richard T.
TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC
BHRICHLIA AND METHODS OF USE
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: HALE AND DORR LLP
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: United States
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,046A
FILING DATE: 24-Apr-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Superko, Colleen
REGISTRATION NUMBER: 39,850
REFERENCE/DOCKET NUMBER: 106,941,155
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 526-6000
TELEFAX: (617) 526-5000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-066-046-4

Query Match Best Local Similarity 4.9%; Score 125.5; DB 4; Length 748; Matches 66; Conservative 33; Mismatches 104; Indels 81; Gaps 12;

QY 147 -NRSPEPARTLG-QGRGCFQGSKAIIKHLKRVVEAGHINTPTGMSPLAAVQANAEASNL 205

QY 341 ---KFEBILKHKHAAARAEAAAGTP-----GPAWDDGAGFTSDFGRCIT 383

QY 206 KEANKIVNFILHKGADLSSTEHTGTPALHLATAANGNHTAMLLIDKGAPATORDARGTA 265

QY 384 LRVAVANGLGNKAKLTKMQAGERAKYDFVEMACPGAGGGQQPRSTDQATQKRQAL 443

QY 266 LHTAAANGDGKLYRMIAK-----KCPDSC----QPLCSDMGDTALHE-AL 305

QY 444 Y--NLDEK---SFLRSRNHNPSTRELVOTYLGPLEGHKAHELLH 482

QY 306 YSDNTERCKFLKMKESRKHLNSSSFFGDLNTPQ_EANGDTLLH 349

Db 1148 ---POQRLLTESWEATERRAGIDP--HSLSGRSGVYAGWMOEQYGRPLAEGE---- 1196

QY 308 FTGTPPLRSL-----SEVRGMDQKIKETNTITWPAPGS----- 340

Db 1197 ---GSDGYLILTGTSGSVWSG--RYAVTLLGEGRPAVTVDTCSSLVVALHQAQLRG 1250

QY 324 MDGIREKNTINWMPGSKFRELKHKHAA---RAEAAGHPTGPGLAWDGAGF----- 373

Db 1251 CDMALAGGYVWAGC-MFVEFSPRGLADGRKAFADG-TWAEAGNVLVERLS 1308

QY 374 -----TSIG-G-RGITL-----RVAVANGLGNKAKLTKMQAGEAK 408

Db 1309 DARRLGHFVLAUVEGSAVNQDAGSNGLTAPSGPSOEV-IRQALGNARLTWADVDEVAH 1367

QY 409 -----YDFEVEMACPGACV--GGQP 428

Db 1368 GTGTRLGLDPIEAQAL-LGTYGRDRGGP 1395

RESULT 2

US-08-804-227C-2

Sequence 2, Application US/08804227C

Patent No. 5076991

GENERAL INFORMATION:

APPLICANT: DeHoff, Bradley S.

APPLICANT: Kuscross, Stuart A.

APPLICANT: Rosleck, Paul R., Jr.

APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501

STREET: LILY CORPORATE CENTER

CITY: INDIANAPOLIS

STATE: IN

ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY disk

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCI(DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804, 227C

APPLICATION NUMBER: US/08/804, 227C

CLASSIFICATION: February 21, 1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 276-2459

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 4472 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-804-227C-2

RESULT 3

US-08-887-534A-47

Sequence 47, Application US/08887534A

Patent No. 6455323

GENERAL INFORMATION:

APPLICANT: Holden, David W.

TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive/6300 Sears Tower

CITY: Chicago

STATE: Illinois

ZIP: 60605

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PATENT Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/887, 534A

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien

REGISTRATION NUMBER: 33,547

REFERENCE/DOCKET NUMBER: 28341/33996

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-6448

TELEX: (312) 474-6600

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 857 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-887-534A-47

Query Match 4.3% Score 112; DB 2; Length 4472;

Best Local Similarity 23.4%; Pred. No. 0.29; Mismatches 91; Conservatism 46; MisMatches 124; Indels 128; Gaps 24;

QY 143 SELHLRHLTEHLP-----HSDERLPMFTSCC--PGWYIAMEKSYPLI--- 185

QY 1032 AQFLH--TELDGSPSSVLAQPKSFEAQEPFLAVVMGCRFPGGVGSPEALWRLVVEGV 1089

QY 186 ---PV-----VSCKSPQMLAMVSKY-----AEKKGIAPKDMVNS 221

QY 1090 DAVSPFPGDGMWVSLIUDFPGVAA--KSVKCGSLHDAEFLAEFGLSPKRAVAMU 1147

QY 222 IMPCTRKQSEADRDPMFCVD---ADPLRQDHVITVLEGNIFKRGINLAELPGEWDN 278

Db 1148 ---POQRLLTESWEATERRAGIDP--HSLSGRSGVYAGWMOEQYGRPLAEGE---- 1196

QY 108 GATTPKQLEGLR-LGFDEVDFDPLGFDLITIMEGSEELHLRLEHLEHPP--SDEPL 163

Db 312 GATTDEYROYIEKDAALERRFQKVFAEPSV-EDTIAILRGLKERYELHHWQITDPAI 370

QY 164 PMFTSCCPGWMAMEKSYP--DILPYVSSCKSPQMLAMVNS 221

QY 199

QY 200 AMVSKYLAEKKGATPKDAMVSMVTPCTRROSEADRWFCVDAD----PTLR-OLDHVIFT 254

Db 429 LMKESDEASKRKL--DMLNEELSDKERQYSELEEFKAESLGTQTKAELQAKTA 485

QY 255 VE---IGNIKERGINLAEELPEGEGWDPMGCVSGGAGLFGT----GGYMEALRTRV 305

Db 486 IEGARRYGDALRMSELQYKPILEKOL----EATOLEGKTMWLLRNKVTDAAEAVL 540

QY 306 ELEFTGTPAPLRLSEVRGMGDKETNITWVAPGSKEFELLKHKRAARAE---- 356

Db 541 ARWTGIVPSRMSE----REKLRM----DOELHHRVIGNEAVDAWSNATR 585

QY 357 -AAHGTGCPPLAWDGAGETSEDGRRGIFTIRVAVANGLGNNAKKLITKMQAGE--AKYDFEV 413

Db 586 RSRAGLADPNRPIGSFLFLGPTVGKTECLKALANFMFDSDDEANVRIDSEFMEKHSVR 645

QY 414 IMACPAGCAGV--GGGGPQRSTDKAITQRQAAVNLDEKSTLRRSH 456

Db 646 LVGAPPGVYVGYEBGGYLTAA----VRRRPYSVILLDE----VERKAH 683

RESULT 4

US-07-642-734C-2

Sequence 2, Application US/07642734C

Patent No. 5824513

GENERAL INFORMATION:

APPLICANT: Katz, L

APPLICANT: Donadio, S

APPLICANT: Mcalpine, J B

TITLE OF INVENTION: Recombinant DNA Method for Producing

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Edward H. Gorman

STREET: Abbott Laboratories D377/AP6D-2 One Abbott

CITY: Park Rd

STATE: Abbott Park

STATE: IL

COUNTRY: US

ZIP: 60064-3500

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/642,734C

FILING DATE: 17-JAN-91

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Danckers, Andreas M

REGISTRATION NUMBER: 32052

REFERENCE/DOCKET NUMBER: 4952.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-9396

TELEFAX: 708-938-2622

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 3491 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-642-734C-2

Query Match 4.1%; Score 105.5; DB 2; Length 3491;

Best Local Similarity 21.6%; Pred. No. 0.93; Matches 117; Conservative 60; Mismatches 213; Indels 151; Gaps 25;

QY 10 AAVSIRGSSCRARQVAPRPLAATVRLVALTAP----RRIGNVACAAAPAAEAPL 65

Db 405 ASTGRHRKTCERLIA----VNGETAAALGRDADAEATFREGLDLSVLAQLEAK-- 456

RESULT 5

US-08-439-009A-2

Sequence 2, Application US/08439009A

Patent No. 6004787

GENERAL INFORMATION:

APPLICANT: Donadio, S

APPLICANT: Katz, L

APPLICANT: Mcalpine, J B

TITLE OF INVENTION: Method of Directing Biosynthesis of

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Steven F. Weinstock

STREET: Abbott Laboratories D377/AP6D-2 One Abbott

CITY: Park Rd

STATE: Abbott Park

STATE: IL

COUNTRY: US

ZIP: 60064-3500

COMPUTER READABLE FORM:

COMPUTER: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/439009A

FILING DATE: 11-MAY-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Casuto, Dianne

REGISTRATION NUMBER: 40,943

REFERENCE/DOCKET NUMBER: 4952.US.D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 847-938-3137

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 3491 amino acids
 TYPE: amino acid
 . TOPOLOGY: linear
 . MOLECULE TYPE: protein

US-0-439-009A-2
 Query Match 4.18; Score 105.5; DB 3; Length 3491;
 Best Local Similarity 21.6%; Pred. No. 0.93; Mismatches 213; Indels 151; Gaps 25;
 Matches 117; Conservative 60; Mismatches 213; Indels 151; Gaps 25;

Qy 10 AAVSIRGSSCRARQVADPRAPLAATVRAVALTEAPA---RRIGWVACAAAPAAEAPL 65
 Db 405 ASPGCTRHRITCERILLA---WVNGETALAGREADATFRELGLDSVLAQLRK--- 456
 Qy 66 SHVQOAL---AELAKPKDPTRKRVHCVQVAVAPAVVIAETG----- 104
 Db 457 --VSAAGIGREVNIAILYDHPTRALAAEALAGTEVQRETRARTNAAAGSPVAVVAMAC 514
 Qy 105 --LAPGATTPKQ---LAEGLRL---GFDEVDTLFGADLTIMEGSELLHLRLTEH 152
 Db 515 RLPGGVSTPEEFWELLSEGIRDVAGLPTDRGMD---LDSLFRDP---RSGP----- 561
 Qy 153 LEAHPHSEDEPLPMFTSCCPGMWAMLERKSYDPLIPYVSSCISQPMMLAAMKSYLAEKKG 212
 Db 562 --AHQRGGGFLHEATAEDPAFGM-----SPREALADPQQRLMELSWEVLERAGI 611
 Qy 213 ARKDMVWSIMPCTRKQSEADRDBWFCCVDAADPTLRLDHVITVVEGLNIFKERGINLAELP 272
 Db 612 PPTSL-----QASP-----TGFVFLGLIPOEYGPRLAE--- 638
 Qy 273 EGEWDNPMPGMGVSSAGVLFGTGTMGVMEALRTAYELFTGTLPLRPLSVE----- 321
 Db 639 -----GEGVEGYLMTGTTSV--ASGRIAYTL--GLRGPATSVDTACSSLVAVHLA 687
 Qy 322 -----RGMDG3IKET-RITMPAPGSKFEELLKHRAARAAAHGTPGPIALWDGGAGFTS 375
 Db 688 COSLRRGESSLAMAGGVYMPMLVDFSRNNSLAPDGRCKASA-----GANGFQ 740
 Qy 376 EDGRGGITLRLAVANGLNNAKKLTMQAGEAKYDFVEI-MACPAGCVGGQQPRSTDKA 434
 Db 741 AEGAGHMLLE-RLSDARRNGHPVAVLRLGTRAVNSDAGSNGLISAP---NGRAQYRVIQIA 795
 Qy 435 ITQK--RQAALYLNDEKSTLRSRNHSPIRELYDY--LGEP--LGHKAHELLHTHYVA 487
 Db 796 LAESCLGPPADIDAVEHGTTGTRLGDPPIEARLFEAYGRDREQPLHGSVSNLGHTOAA 855
 Qy 488 G 488
 Db 856 G 856

RESULT 6
 US-0-036-987A-4
 ; Sequence 4, Application US/09036987A
 ; Patent No. 6143526
 ; GENERAL INFORMATION:
 ; APPLICANT: Baltz, Richard H.
 ; APPLICANT: Broughton, Mary C.
 ; APPLICANT: Crawford, Kathryn P.
 ; APPLICANT: Madduri, Krishnamurthy
 ; APPLICANT: Merlo, Donald J.
 ; APPLICANT: Treadaway, Patti J.
 ; APPLICANT: Turner, Jan R.
 ; APPLICANT: Waldon, Clive
 ; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
 ; NUMBER OF SEQUENCES: 39
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dow AgroSciences LLC Patent Department
 ; STREET: 9330 Zionsville Road
 ; CITY: Indianapolis
 ; STATE: Indiana

COUNTRY: USA
 ZIP: 46268
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/036, 987A
 FILING DATE: 09-MAR-1998
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Stuart, Donald R
 REGISTRATION NUMBER: 28,479
 REFERENCE/DOCKET NUMBER: 50,608
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (317)337-8816
 TELEFAX: (317)337-4847
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3170 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-09-036 987A-4
 Query Match 3.9%; Score 100.5; DB 4; Length 3170;
 Best Local Similarity 21.3%; Pred. No. 2.6; Mismatches 190; Indels 175; Gaps 33;
 Matches 116; Conservative 64; Mismatches 190; Indels 175; Gaps 33;
 Qy 43 EPARRLGNVACAAAPAAEAPLWVQVQALAEELAKPKDPTRK---VGVQVAVAVR 97
 Db 1433 EVPAARKAMPANGPRPEGG---SPFARNLAEELPEAQ---RHEELVULVCAQATV--- 1481
 Qy 98 AIAETFLGLPAGTPEKQLAEGLRIGFDEVDTLFGADL-----TIME-- 140
 Db 1482 -----LGHSGRVEQPERAFLRG---FDSIAMDVRNRLTTAIGRLRPLTTVFDYP 1530
 Qy 141 EGSEHLHLRLTEHL-----EAHPHSDEPLPMFTSC----- 170
 Db 1531 NPAALAAHLEELYVGDVASAATVASAPASDEPIAVAMSCREFGGAHSPEDLWRLVAG 1590
 Qy 171 -----PGWIAHMSY-POLIPVSSCKSPQMLAAM---VKSIAEKKGIA 213
 Db 1591 TEVIGEFPSPDRGMDA---EGLYDPD---ASRPTTYARMAGFLYDAGEADLGPS 1641
 Qy 214 PKDAMVWSIMPCTRKQSEADRDBWFCCVDAADPTLRLDHVITVVEGLNIFKERG--INL 268
 Db 1642 PREALMD---POORLYLEIAEALERAGIDPLSLKSGVGYYGAGS---RGVATDV 1693
 Qy 269 AELPGEWNPMPGMGVSSAGVLFGTGTMGVMEALRTAYEL-FGCTPLP-----R 315
 Db 1594 RQFPE-EAEGYLTGTSASVLSG-----RVASFGFEGPANTVDTACSSLVVALH 1742
 Qy 316 LSISVERG---MDG3IKETNITMVAPGSKFEELLKHRAA---RAEAAAHGTPGPIALWDGG 370
 Db 1743 LACOSLRSGECDCIALLAGGVTMSTP-EMFVEPSRQRGLAPDGRCKSFAESADG-TGSGEG 1800
 Qy 371 AGFISDEGRGGITLRLAVANGLNNAKKLTMQAGEAKYDFVEI-MACPAGCVGGQQPR 429
 Db 1801 A-----GLLLERLSLDAHNRGHRVAVYRGAVNODGASNGLAAP---NGSQQR 1847
 Qy 430 STDKAITOKRQAALYLNDEKSTLRSRNHSPIRELYDYLGEPGLGKAHELLHTHYVAGG 489
 Db 1848 VTNQALA---NAAALSAQSDAV--EAGT-----GTRLGDP---EAQALIATY--GO 1890
 Qy 490 VEERD 494
 Db 1891 ARERD 1895

RESULT 7
 US-09-370-700-4

Sequence 4, Application US/09370700
 Patent No. 624350
 GENERAL INFORMATION:
 APPLICANT: Baltz, Richard H
 APPLICANT: Broughton, Mary C
 APPLICANT: Crawford, Kathryn P
 APPLICANT: Madduri, Krishnamurthy
 APPLICANT: Treadaway, Patti J
 APPLICANT: Waldron, Clive
 TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
 FILE REFERENCE: 50489 DIV1
 CURRENT APPLICATION NUMBER: US/09/370,700
 CURRENT FILING DATE: 1999-06-09
 EARLIER APPLICATION NUMBER: US 09/36987
 EARLIER FILING DATE: 1998-03-09
 NUMBER OF SEQ ID NOS: 39
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 4
 LENGTH: 3170
 TYPE: PRT
 ORGANISM: *Saccharopolyspora spinosa*
 US-09-370-700-4

Query Match 3.9%; Score 100.5; DB 4; Length 3170;
 Best Local Similarity 21.3%; Pred. No. 2.6; Mismatches 190; Gaps 33;
 Matches 116; Conservative 64; Indels 175; Gaps 33;
 Qy 43 EAPARRIGGNVACAAAPAAEAPLSHVOQALABLAQPKDPTRKH---VCVQAVAPRV 97
 Db 1433 EVAAPKAMPANGPAEFGG---SPFARNLSELPEAQ---RRHELVLDLVCRAQVATV-- 1481
 Qy 98 AIAETLGLAP-GATTPKQLAEGURRLGTDDEVFTLFRADL-----TIME-- 140
 Db 1482 ---LGHGSREEVOPERAFLAG---FDSIMAVDLRNLTTATGLRLIPTTVFEDYP 1530
 Qy 141 EGSELLHLRFLTEIL---EAMPHSDPPLPMETSC----- 170
 Db 1531 NPAALAAHLEELVGDVASAATASAPASDEPIATVMSCRPGGAHSPELDWLRLVAAG 1590
 Qy 171 -----POWIAMLEKSY-PDLIPIVSSCKSPQMLAAM-----VKSylaEKKGIA 213
 Db 1591 TEVIGEEFFSDRGWDA--EGLYQPD---ASRPGTYTARMAGFLDADGEEDADLFGIS 1641
 Qy 214 PKDMVMSIMPCTRKQSEBADRQFCVD---ADPTLROLDHVITVLEGNIFKERRG--INL 268
 Db 1642 PREALAMD---POORLWLEFAWEALEFAGIDPLSLKGSGVTVIGAGS---RGATDV 1693
 Qy 269 AELPEGEFTDNPMLGVGSGAGVLRGTTGWMEEALRTAVEL-FGTCPLP-----R 315
 Db 1694 ROPPE-EAGYLLTGTSASVLSG-----RVAYSGFEGCPAVTVDTCSSSLVALH 1742
 Qy 316 LSUSEVRG--MDGKETNTWVPPGSKFEELKHKRA--RAEAAGHGPGLAWDGG 370
 Db 1743 LACQSLSSGECOLALAGGTVWSTP-EMFVEVSRQGIAAPDCKSSEAESADG-TGWEG 1800
 Qy 39 LATELAPARRIGGNVACAAAPAAEAPLSHVOQALABLAQPKDPTRKH---VCVQAVAPRV 98
 Db 1 MATAAAAATSTALTTGATAAPKARRHLATRRAIA-----PIR---CSASPAMPMA 50
 Qy 99 IAEETLGLAP-GATTPKQ---LAEGLRLGFDDEVFTLFGADLTMEGESELHR---LT 150
 Db 51 PRPT-PLRPGWGPDRPKGADLIVESLERCGYRDVFAYPGGSMET---HQALTRSVIA 105
 Qy 151 EHLEAHPHSDEPLPMETSCCPGWIAMEKSY-PDLIPIVSSCKSPQMLAAMVKSylaEKKGIA 209
 Db 106 NHFRHEOGE-----AFASAGYARSSGRGVCVTATSGPATNLVSLDAL 151
 Qy 210 -----KGIAPKDMV-----MVSIMPCTRKQSEBADRQFCVDADPTLQDLDHVIT 254
 Db 152 LDSPVMVWALTQVPRMIGTDAFOETPIVEVTR--SITKHNLVLDVDTIPVVAEFL 209
 Qy 255 VELGNIFKERRGIAALPEG-----EDWDPMGVOSGAGVLFGTGGMVMEALRVAE 306
 Db 210 AASSG---RPGPVLYDIPKDTQOQMVAYPWDRPMSL-----PGYIARLUPKPPATE 255
 Qy 307 LFTGTPPLPLRLSISEVERGMDGIGIETNINW---PAPGSKFEELKHKRAARAAEAMGTP 362
 Db 256 L-----LEQVRLRVLGESRRPVLVYGGCAASG---EEL-----RREVLTGIP 295

RESULT 8
 US-08-426-125-4
 ; Sequence 4, Application US/09426125
 ; Patent No. 5853973

RESULT 9
US-08-455-355-4

Sequence 4, Application US/08455355

Patent No. 5928937

GENERAL INFORMATION:

APPLICANT: Kakefuda, Genichii
APPLICANT: Ott, Karl Heinz
APPLICANT: Kwagon, Jae-Gyu

TITLE OF INVENTION: Structure-Based Designed Herbicide

TITLE OF INVENTION: Resistant Products

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-7513

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patient Release #1.0, version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/455, 355

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Robinson, Joseph
REGISTRATION NUMBER: 33, 448
REFERENCE/DOCKET NUMBER: 0646/1A674-US1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)-527-7783
TELEX: 236687

SEQUENCE CHARACTERISTICS:

LENGTH: 638 amino acids

TYPE: amino acid

STRANDEDNESS: linear

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: zea mays

US-08-455-355-4

Query Match 3.9%; Score 99.5; DB 2; Length 638;
Best Local Similarity 22.3%; Pred. No. 0.23; Gaps 23;
Matches 113; Conservative 56; Mismatches 219; Indels 119; Gaps 23;

QY 39 LATLEAPARRGNVACAAAPAAAPLSIVQALAEALKPKDUDPTRKHCVQVAPAVVA.98

Db 1 MATAAAASTALTGATAAATKARRAHLLTRALAA-----PIR----CSAASPAMPA 50

Db 99 IAETVGLAP-GATTPKQ---LAEGLRIGFDEVDFDTLEGADLUTIMEEGSELUR--LT 150

QY -.51 PPAT-PLRKGKPTPKRGADLIVESLERGVROFVAYPGASMEI---HQALTRSPVIA 105

Db 151 EHLEAHPHSDEPLPMTSCPGWIMLERSYDPLIPVYSSCKSPOMMLAMVSKYLABR- 209

RESULT 10
US-08-104-227C-10

Sequence 10, Application US/08804227C

Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.

APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Thomas G. Plant 1501
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII(DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/104-227C

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35, 784

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459

SEQUENCE CHARACTERISTICS:

LENGTH: 3724 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-104-227C-10

Query Match 3.9%; Score 99.5; DB 2; Length 3724;
Best Local Similarity 22.8%; Pred. No. 4.4; Gaps 22;
Matches 124; Conservative 54; Mismatches 255; Indels 111; Gaps 22;

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: ASCIDOS) Text only
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/804-227C
 FILING DATE: February 21, 1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Plant, Thomas, G.
 REGISTRATION NUMBER: 35,784
 REFERENCE/DOCKET NUMBER: X-8231
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-276-2459
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4550 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 US-08-804-227C-8

Query Match

Best local similarity 3.9%; Score 99.5; DB 2; Length 4550;

Matches 114; Conservative 62; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAKPKDPP----- 82

Db 279 VLRACAOARVDPAEVREVELHGTGTPVGDPEVHALGVHGSGRPADDPLVGSVKNI 338

QY 83 -----TRKHVCV-----QVAPAVRVAIAETGLARGA 109

Db 339 GHLEGAAGTAGLVKAALCLERTLPGSLNFATPSPAIDLQRLKVQTAEE-LPLAPGG 397

QY 110 TPPKQLAEDLRLRGDFEVDTLFGADLTIMEEGSELLHLTELEAHHSDEPLPMFTSC 169

Db 398 A---PLLAGVSSSGI-----GTCNCHVLEHLPSRPP----- 427

QY 170 CGPWIAMLKSYPPDILPVYSSCKSPQMMIAAMVK-SYLAEKKGIAPKDMVM----- 220

Db 428 ---AVSVAAASLPDPVPLLISARSEGALKHQAVRLGEYERVERGADPROVAYSLASTRLF 483

QY 221 --SIMPCRKQS-----EADR-----DW----- 236

Db 484 EHRAWVPCGGRGELVAALGGFAAGRVGGVRSGRGAVPGSGVGLFTGOGQWVGNGRLYA 543

QY 237 ---FCVDAADPTLROLDHV---ITVVEGNIFVERGINLAELPEGEWNINPMVGSGAGV 288

Db 544 GGGVFAEVDEVLNSGEVGDGRSLRDVMDGVDVDAAGA-----DAGAGAGAGVS 595

QY 289 LGFTGTVGV-----MEALRTAYE-----LFTGTPPLPLSSEYRGMGJTK 328

Db 596 GSGSVVGLLGRTEFAQPALEVALFRALEARGVESVVLGHSGVGEVAAYAVGLSLG 655

QY 329 ETNTIWMPPAPGSKF-----EELKKHARAEE---AAHGTCPOLAND 369

Db 656 DA-VRLVWVARGLMGLPVGGMNSVGASESWSRGVVEGLGEWVSAAVING-PRSVWLSG 713

QY 370 GAG-----FTSEDGRGGITLKVAYANG-----LGNAAKLTTKMQAGEAKYDFVETM 415

Db 714 DVGVLESVRASLMDGVEVERRLDVSHGFSVSLMPEVPLVGLGFRGVVESLBRGRVRGVWWS 773

QY 416 ACAPACVGCG---GOP 428

Db 774 GVSGGVVGSGELGDP 788

US-08-804-198-2

Query Match

Best local similarity 18.5%; Pred. No. 6.2; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAKPKDPP----- 82

Db 279 VLRACAOARVDPAEVREVELHGTGTPVGDPEVHALGVHGSGRPADDPLVGSVKNI 338

QY 83 -----TRKHVCV-----QVAPAVRVAIAETGLARGA 109

Db 339 GHLEGAAGTAGLVKAALCLERTLPGSLNFATPSPAIDLQRLKVQTAEE-LPLAPGG 397

QY 110 TPPKQLAEDLRLRGDFEVDTLFGADLTIMEEGSELLHLTELEAHHSDEPLPMFTSC 169

Db 398 A---PLLAGVSSSGI-----GTCNCHVLEHLPSRPP----- 427

QY 170 CGPWIAMLKSYPPDILPVYSSCKSPQMMIAAMVK-SYLAEKKGIAPKDMVM----- 220

Db 428 ---AVSVAAASLPDPVPLLISARSEGALKHQAVRLGEYERVERGADPROVAYSLASTRLF 483

QY 221 --SIMPCRKQS-----EADR-----DW----- 236

Db 484 EHRAWVPCGGRGELVAALGGFAAGRVGGVRSGRGAVPGSGVGLFTGOGQWVGNGRLYA 543

QY 237 ---FCVDAADPTLROLDHV---ITVVEGNIFVERGINLAELPEGEWNINPMVGSGAGV 288

Db 544 GGGVFAEVDEVLNSGEVGDGRSLRDVMDGVDVDAAGA-----DAGAGAGAGVS 595

US-08-804-198-3

Query Match

Best local similarity 18.5%; Pred. No. 6.2; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAKPKDPP----- 82

Db 279 VLRACAOARVDPAEVREVELHGTGTPVGDPEVHALGVHGSGRPADDPLVGSVKNI 338

QY 83 -----TRKHVCV-----QVAPAVRVAIAETGLARGA 109

Db 339 GHLEGAAGTAGLVKAALCLERTLPGSLNFATPSPAIDLQRLKVQTAEE-LPLAPGG 397

QY 110 TPPKQLAEDLRLRGDFEVDTLFGADLTIMEEGSELLHLTELEAHHSDEPLPMFTSC 169

Db 398 A---PLLAGVSSSGI-----GTCNCHVLEHLPSRPP----- 427

QY 170 CGPWIAMLKSYPPDILPVYSSCKSPQMMIAAMVK-SYLAEKKGIAPKDMVM----- 220

Db 428 ---AVSVAAASLPDPVPLLISARSEGALKHQAVRLGEYERVERGADPROVAYSLASTRLF 483

QY 221 --SIMPCRKQS-----EADR-----DW----- 236

Db 484 EHRAWVPCGGRGELVAALGGFAAGRVGGVRSGRGAVPGSGVGLFTGOGQWVGNGRLYA 543

QY 237 ---FCVDAADPTLROLDHV---ITVVEGNIFVERGINLAELPEGEWNINPMVGSGAGV 288

Db 544 GGGVFAEVDEVLNSGEVGDGRSLRDVMDGVDVDAAGA-----DAGAGAGAGVS 595

US-08-804-198-4

Query Match

Best local similarity 18.5%; Pred. No. 6.2; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAKPKDPP----- 82

Db 279 VLRACAOARVDPAEVREVELHGTGTPVGDPEVHALGVHGSGRPADDPLVGSVKNI 338

QY 83 -----TRKHVCV-----QVAPAVRVAIAETGLARGA 109

Db 339 GHLEGAAGTAGLVKAALCLERTLPGSLNFATPSPAIDLQRLKVQTAEE-LPLAPGG 397

QY 110 TPPKQLAEDLRLRGDFEVDTLFGADLTIMEEGSELLHLTELEAHHSDEPLPMFTSC 169

Db 398 A---PLLAGVSSSGI-----GTCNCHVLEHLPSRPP----- 427

QY 170 CGPWIAMLKSYPPDILPVYSSCKSPQMMIAAMVK-SYLAEKKGIAPKDMVM----- 220

Db 428 ---AVSVAAASLPDPVPLLISARSEGALKHQAVRLGEYERVERGADPROVAYSLASTRLF 483

QY 221 --SIMPCRKQS-----EADR-----DW----- 236

Db 484 EHRAWVPCGGRGELVAALGGFAAGRVGGVRSGRGAVPGSGVGLFTGOGQWVGNGRLYA 543

QY 237 ---FCVDAADPTLROLDHV---ITVVEGNIFVERGINLAELPEGEWNINPMVGSGAGV 288

Db 544 GGGVFAEVDEVLNSGEVGDGRSLRDVMDGVDVDAAGA-----DAGAGAGAGVS 595

US-08-804-198-5

Query Match

Best local similarity 18.5%; Pred. No. 6.2; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAKPKDPP----- 82

Db 279 VLRACAOARVDPAEVREVELHGTGTPVGDPEVHALGVHGSGRPADDPLVGSVKNI 338

QY 83 -----TRKHVCV-----QVAPAVRVAIAETGLARGA 109

Db 339 GHLEGAAGTAGLVKAALCLERTLPGSLNFATPSPAIDLQRLKVQTAEE-LPLAPGG 397

QY 110 TPPKQLAEDLRLRGDFEVDTLFGADLTIMEEGSELLHLTELEAHHSDEPLPMFTSC 169

Db 398 A---PLLAGVSSSGI-----GTCNCHVLEHLPSRPP----- 427

QY 170 CGPWIAMLKSYPPDILPVYSSCKSPQMMIAAMVK-SYLAEKKGIAPKDMVM----- 220

Db 428 ---AVSVAAASLPDPVPLLISARSEGALKHQAVRLGEYERVERGADPROVAYSLASTRLF 483

QY 221 --SIMPCRKQS-----EADR-----DW----- 236

Db 484 EHRAWVPCGGRGELVAALGGFAAGRVGGVRSGRGAVPGSGVGLFTGOGQWVGNGRLYA 543

QY 237 ---FCVDAADPTLROLDHV---ITVVEGNIFVERGINLAELPEGEWNINPMVGSGAGV 288

Db 544 GGGVFAEVDEVLNSGEVGDGRSLRDVMDGVDVDAAGA-----DAGAGAGAGVS 595

US-08-804-198-6

Query Match

Best local similarity 18.5%; Pred. No. 6.2; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAKPKDPP----- 82

Db 279 VLRACAOARVDPAEVREVELHGTGTPVGDPEVHALGVHGSGRPADDPLVGSVKNI 338

QY 83 -----TRKHVCV-----QVAPAVRVAIAETGLARGA 109

Db 339 GHLEGAAGTAGLVKAALCLERTLPGSLNFATPSPAIDLQRLKVQTAEE-LPLAPGG 397

QY 110 TPPKQLAEDLRLRGDFEVDTLFGADLTIMEEGSELLHLTELEAHHSDEPLPMFTSC 169

Db 398 A---PLLAGVSSSGI-----GTCNCHVLEHLPSRPP----- 427

QY 170 CGPWIAMLKSYPPDILPVYSSCKSPQMMIAAMVK-SYLAEKKGIAPKDMVM----- 220

Db 428 ---AVSVAAASLPDPVPLLISARSEGALKHQAVRLGEYERVERGADPROVAYSLASTRLF 483

QY 221 --SIMPCRKQS-----EADR-----DW----- 236

Db 484 EHRAWVPCGGRGELVAALGGFAAGRVGGVRSGRGAVPGSGVGLFTGOGQWVGNGRLYA 543

QY 237 ---FCVDAADPTLROLDHV---ITVVEGNIFVERGINLAELPEGEWNINPMVGSGAGV 288

Db 544 GGGVFAEVDEVLNSGEVGDGRSLRDVMDGVDVDAAGA-----DAGAGAGAGVS 595

US-08-804-198-7

Query Match

Best local similarity 18.5%; Pred. No. 6.2; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAKPKDPP----- 82

Db 279 VLRACAOARVDPAEVREVELHGTGTPVGDPEVHALGVHGSGRPADDPLVGSVKNI 338

QY 83 -----TRKHVCV-----QVAPAVRVAIAETGLARGA 109

Db 339 GHLEGAAGTAGLVKAALCLERTLPGSLNFATPSPAIDLQRLKVQTAEE-LPLAPGG 397

QY 110 TPPKQLAEDLRLRGDFEVDTLFGADLTIMEEGSELLHLTELEAHHSDEPLPMFTSC 169

Db 398 A---PLLAGVSSSGI-----GTCNCHVLEHLPSRPP----- 427

QY 170 CGPWIAMLKSYPPDILPVYSSCKSPQMMIAAMVK-SYLAEKKGIAPKDMVM----- 220

Db 428 ---AVSVAAASLPDPVPLLISARSEGALKHQAVRLGEYERVERGADPROVAYSLASTRLF 483

QY 221 --SIMPCRKQS-----EADR-----DW----- 236

Db 484 EHRAWVPCGGRGELVAALGGFAAGRVGGVRSGRGAVPGSGVGLFTGOGQWVGNGRLYA 543

QY 237 ---FCVDAADPTLROLDHV---ITVVEGNIFVERGINLAELPEGEWNINPMVGSGAGV 288

Db 544 GGGVFAEVDEVLNSGEVGDGRSLRDVMDGVDVDAAGA-----DAGAGAGAGVS 595

US-08-804-198-8

Query Match

Best local similarity 18.5%; Pred. No. 6.2; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAKPKDPP----- 82

Db 279 VLRACAOARVDPAEVREVELHGTGTPVGDPEVHALGVHGSGRPADDPLVGSVKNI 338

QY 83 -----TRKHVCV-----QVAPAVRVAIAETGLARGA 109

Db 339 GHLEGAAGTAGLVKAALCLERTLPGSLNFATPSPAIDLQRLKVQTAEE-LPLAPGG 397

QY 110 TPPKQLAEDLRLRGDFEVDTLFGADLTIMEEGSELLHLTELEAHHSDEPLPMFTSC 169

Db 398 A---PLLAGVSSSGI-----GTCNCHVLEHLPSRPP----- 427

QY 170 CGPWIAMLKSYPPDILPVYSSCKSPQMMIAAMVK-SYLAEKKGIAPKDMVM----- 220

Db 428 ---AVSVAAASLPDPVPLLISARSEGALKHQAVRLGEYERVERGADPROVAYSLASTRLF 483

QY 221 --SIMPCRKQS-----EADR-----DW----- 236

Db 484 EHRAWVPCGGRGELVAALGGFAAGRVGGVRSGRGAVPGSGVGLFTGOGQWVGNGRLYA 543

QY 237 ---FCVDAADPTLROLDHV---ITVVEGNIFVERGINLAELPEGEWNINPMVGSGAGV 288

Db 544 GGGVFAEVDEVLNSGEVGDGRSLRDVMDGVDVDAAGA-----DAGAGAGAGVS 595

US-08-804-198-9

Query Match

Best local similarity 18.5%; Pred. No. 6.2; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAKPKDPP----- 82

Db 279 VLRACAOARVDPAEVREVELHGTGTPVGDPEVHALGVHGSGRPADDPLVGSVKNI 338

QY 83 -----TRKHVCV-----QVAPAVRVAIAETGLARGA 109

Db 339 GHLEGAAGTAGLVKAALCLERTLPGSLNFATPSPAIDLQRLKVQTAEE-LPLAPGG 397

QY 110 TPPKQLAEDLRLRGDFEVDTLFGADLTIMEEGSELLHLTELEAHHSDEPLPMFTSC 169

Db 398 A---PLLAGVSSSGI-----GTCNCHVLEHLPSRPP----- 427

QY 170 CGPWIAMLKSYPPDILPVYSSCKSPQMMIAAMVK-SYLAEKKGIAPKDMVM----- 220

Db 428 ---AVSVAAASLPDPVPLLISARSEGALKHQAVRLGEYERVERGADPROVAYSLASTRLF 483

QY 221 --SIMPCRKQS-----EADR-----DW----- 236

Db 484 EHRAWVPCGGRGELVAALGGFAAGRVGGVRSGRGAVPGSGVGLFTGOGQWVGNGRLYA 543

QY 237 ---FCVDAADPTLROLDHV---ITVVEGNIFVERGINLAELPEGEWNINPMVGSGAGV 288

Db 544 GGGVFAEVDEVLNSGEVGDGRSLRDVMDGVDVDAAGA-----DAGAGAGAGVS 595

US-08-804-198-10

Query Match

Best local similarity 18.5%; Pred. No. 6.2; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAKPKDPP----- 82

Db 279 VLRACAOARVDPAEVREVELHGTGTPVGDPEVHALGVHGSGRPADDPLVGSVKNI 338

QY 83 -----TRKHVCV-----QVAPAVRVAIAETGLARGA 109

Db 339 GHLEGAAGTAGLVKAALCLERTLPGSLNFATPSPAIDLQRLKVQTAEE-LPLAPGG 397

QY 110 TPPKQLAEDLRLRGDFEVDTLFGADLTIMEEGSELLHLTELEAHHSDEPLPMFTSC 169

Db 398 A---PLLAGVSSSGI-----GTCNCHVLEHLPSRPP----- 427

QY 170 CGPWIAMLKSYPPDILPVYSSCKSPQMMIAAMVK-SYLAEKKGIAPKDMVM----- 220

Db 428 ---AVSVAAASLPDPVPLLISARSEGALKHQAVRLGEYERVERGADPROVAYSLASTRLF 483

QY 221 --SIMPCRKQS-----EADR-----DW----- 236

Db 484 EHRAWVPCGGRGELVAALGGFAAGRVGGVRSGRGAVPGSGVGLFTGOGQWVGNGRLYA 543

QY 237 ---FCVDAADPTLROLDHV---ITVVEGNIFVERGINLAELPEGEWNINPMVGSGAGV 288

Db 544 GGGVFAEVDEVLNSGEVGDGRSLRDVMDGVDVDAAGA-----DAGAGAGAGVS 595

US-08-804-198-11

Query Match

Best local similarity 18.5%; Pred. No. 6.2; Mismatches 192; Indels 247; Gaps 27;

QY 2 SALVYKPCAAVSISSGSS--CRAROWA-----PRAPLAASIVRVALATEPARR--- 48

Db 233 AAVVILPLADALADGDPVYCVVRVAVGNGGGP-----GLTADREGEA 278

QY 49 LGNYVACAAA-----APAAEAPLHVQOALAEAK

US-08-804-227C-9

Sequence 9, Application US/08804227C
; Patent No. 5876991

GENERAL INFORMATION:

APPLICANT: DeHoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKENIDE SYNTHASE GENES

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA

ZIP: 46265

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII (DOS) Text only

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/804-227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-276-2459

SEQUENCE CHARACTERISTICS:

SEQUENCE ID: 9
SEQUENCE FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1996 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide

Query Match 3.8%; Score 98.5; DB 4; Length 823;
Best Local Similarity 20.9%; Pred. No. 0.44; DB 4; Length 823;
Matches 81; Conservative 53; Mismatches 178; Indels 75; Gaps 16;

QY 108 GATTGPKOLAEGGLRRLGDFDEVDTLF--GADITIMEGSELLHRLHELEAHPH--SDE 161
Db 319 GATT--LDEYRKNIERDAALERRFQPOIQDPTVDETDIEILKGLDRYEAHRINISDE 375

QY 162 PLPMFTSCCPGMIA--MLEKSPVLDLPIVSSCKSPQMLAANVSYLAEKKGIAPKDMV 219
Db 376 ALEAAKLSDRYVSDRFLPKDADLDEAASS----KVALKSHH---- 415

QY 220 VSIMPCTRKOSADRDOWFCVADPTLRQDHLVTTVGLNIT--FKERGINLA---ELPEG 274
Db 416 ---PSNLKEIKQ----EIDKVKNEDAAVHQAQFENAANLRDKQSKLERQYEDAKN 464

QY 275 EWDNPMLGVGSGAGVLCITGGIMEALRTAYEFTGTPPLRPLSLEVRGMDDIKETNTM 334
Db 465 EWKNAQG----GLDTALSEENIAAEVLAGWTGIPILTKINETEDRLLNLEDLHKR 515

QY 335 VPAPGSKFEELKHKRAAAREAAHGPGLAWDGGAGFTSEDGRGGITLRVAVANGIGN 394
Db 516 VIGQNDAVNSISK--AVTRARAGLKDPRK--GSFLGFTGVGKTELARALAESMFG 570

QY 395 AKKLITRNMQAE--AKYDFVETMACPAGCVG--GGGPRSTDKATOKRQAAALYNDERS 450
Db 571 EDDAMIVDMSPFEMKEHAVSLVYAGPFCYVGHDDGQ----LITERKVRKPYSVILFD 623

QY 451 TURRSHE--NPSIRELYDTYLGEPPLG 474
Db 624 EIEKAHPDVFNLUQVLDGHLTDKG 650

Query Match 3.7%; Score 94.5; DB 2; Length 1996;
Best Local Similarity 23.6%; Pred. No. 5.2; DB 2; Length 1996;
Matches 96; Conservative 41; Mismatches 172; Indels 97; Gaps 21;

QY 31 AASTVVRVALATEPAPRRLGNVACAAAPAAEADPLSHVQOQALAKPKDDPTRKHVCQ 90
Db 909 ALAVVFTRDRAGLDTTILHGSGRDRDPT--YFQHDXWYLAQPSRRTDGDLSAAGLE 966

QY 91 VAPAVRVAATAETLGLAGATTPKOLAEGLRLRGDFEVEF--TFLGADL--TMRGSEL- 145
Db 967 VEHPLTTAAVEVLPGTDTEWVGRISAADLFWLWDHILVMDRGVVGTTALLETLVQVGSIG 1026

QY 146 LKRTHELEAHP--HSQDPLM---FTSCCGWIALE-KYDPLIPIVSSCKSPM 195
Db 1027 LPRVAAELVLTPLWTSDRPLQRVVTTAATPGARELTLSRPEPVAAASSSPSP-- 1084

QY 197 MAAVVKSYLAEKKGIAPKDMVMSIMPCTRKOSADRDOWFCVADPTLRQDHLVTTW 256
Db 1085 -----ASPHLT-----AQSDDW-----TRHA 1103

QY 257 LGNIFKERSI--NLAEELPGCEWDNPMSGV--SGAGVLFGTGCGMEA--LRTAY-- 305
Db 1104 SGLLAAGAADDFAEL-TGAWP-PVGAAPLGLAQYPLFAAGAVRYEGAFGLRGLAARR 1161

QY 305 --ELFTGTPPLRSLSEVSGMD---GKETNTWPAVGSKFEELKHKRAAAREAAH 359
Db 1162 GDE/FADVRLP----DAHADVADRYGVHPLAIDAVLHPTASLDP-----GDGHH 1207

QY 360 GTPGPLAWDGGAGFTSEDGRGGITLRVAVANGLGNNAKLTQMAQ 405
Db 1208 GLL-PFSEWTDQG---HGAGHHLRVYAAVQDGAVSTAADIG 1248

Wed Jun 4 16:29:41 2003

Search completed: June 3, 2003, 16:09:43
Job time : 32 secs

cac80065.rai

Page 10

```

Sequence 87, App1
Sequence 2, App1
Sequence 6032, App1
Sequence 18, App1
Sequence 309, App1
Sequence 66, App1
Sequence 5, App1
Sequence 43, App1
Sequence 30, App1
Sequence 178, App1
Sequence 4, App1
Sequence 4, App1
Sequence 4, App1
Sequence 4, App1
Sequence 31, App1
Sequence 6, App1
Sequence 6, App1
Sequence 6, App1
Sequence 22, App1
Sequence 250, App1

```

Db 4600 BEPLALAGL-RVTDLPGATEPAK-----VLDLTLHEVAGRDGMHATL----- 4641
 Qy 265 INLAELPEGEWNDNPMVGVGAGVLFGTGGVMEALRPAELFTGTPURLSLEVRGMD 325
 Db 4642 LGADDLFEDE-----TVALADBLRRT-LEAMAAAPDDRLDRIEVSPG 4684
 Qy 326 G---IKENNITWVAPGSKFEELKHRRAARAAEAAHGTGPLAWDGAGFTSEDGRGG 381
 Db 4685 ERSULVVEENDTARPVVESEVPALFAEQVAAADAVVGEV-VSNT---YRELDARSD 4739
 Qy 382 ITLRVAVANGLGNNAKLLTRMQ-----AG-----EAKYDFVETMA 416
 Db 4740 ALARSIVAVAGVGVESPVVALERSPEVLSAFLAVAKAGGVFVFDLWSWQARYDAV-VAD 4798
 Qy 417 CPAG-----CVGGGQOPRSKDATQKRO-----ALYNYDEKSTL 452
 Db 4799 CGARIAVAVRPMGSLTVVSAGLGGSAAVSGDLTADRAVVLPAQPVPGAVYRMYTSGT 4858
 Qy 453 RR-----SHEN-----PSIRELYD----- 466
 Db 4859 GRPKGVUTTHONLYDLATDRCWGTTPRVLPHAFDASSYEIKVPLNGGTVVWAPRRS 4918
 Qy 467 ---TVLGEPILGHKAELLHLYHVAG 488
 Db 4919 IDATVLRDLIG--AHELTIVHVHTAG 4941

RESULT 2
 US-09-737-149-4
 ; Sequence 4, Application US/09737149

GENERAL INFORMATION:
 ; Patent No. US2003007466A1
 ; APPLICANT: Spaderna, Steven K
 ; APPLICANT: Quinn, Kerry E.
 ; APPLICANT: Shinkets, Richard A.
 ; APPLICANT: Muralidhara, Padigaru
 ; APPLICANT: Spytek, Kimberly A.
 ; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 15966-620 CIP
 ; CURRENT APPLICATION NUMBER: US/09/737,149
 ; CURRENT FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: 60/170,564
 ; PRIOR FILING DATE: 1999-12-14
 ; PRIOR APPLICATION NUMBER: 60/173,165
 ; PRIOR FILING DATE: 1999-11-27
 ; PRIOR APPLICATION NUMBER: 60/173,362
 ; PRIOR FILING DATE: 1999-12-27
 ; PRIOR APPLICATION NUMBER: 60/173,544
 ; PRIOR FILING DATE: 1999-12-29
 ; PRIOR APPLICATION NUMBER: 60/174,404
 ; PRIOR FILING DATE: 2000-01-04
 ; PRIOR APPLICATION NUMBER: 60/174,962
 ; PRIOR FILING DATE: 2000-01-07
 ; PRIOR APPLICATION NUMBER: 60/223,929
 ; PRIOR FILING DATE: 2000-08-09
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 965
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-737-149-4

Query Match 4.7%; Score 122; DB 10; Length 965;
 Best Local Similarity 21.9%; Pred. No. 0.1; Mismatches 120; Conservative 46; Indels 189; Gaps 22; Matches 120; Organism 46; Mismatches 189; Indels 192; Gaps 22;

Db 85 ASLURGLCQALVPPGVAALLAFAPEARPELQLQHFLAMATETPVLSLURREARRAPASAPNP 144
 Qy 113 KQLAEGLRRLRGEEVEDFLGADLTIMEGESELHLRTEHLRPHISDEPLMFTSCPG 172
 Db 145 FHL---QLHWASPLT-----LLDVVLVAVLQGAHAWEDVGLALCRTCQDPG 185
 Qy 173 WIAMEKSY ---PDLLPVSSCKSPOMLAMAKVSKLAEGKJIAKDMWVMSIMCPTRK 228
 Db 186 GLVALWTSRAGRPQQLVLDLSRDTGDAGLRNLAPMAAPVGEAP----- 231
 Qy 229 QSEARDMFCVDAADPTLROLDHVITVLELG-NIFKRCINLAEPEGEWNDNPMVGVSAG 287
 Db 232 -----VPAAVLJGCDIARARRYLEAVPPGPW----- 258
 Qy 288 VLRFGTGGVMEALRATVLELFTGTPPLPRLSLEVR-----GMGDK 328
 Db 259 -LGGTP--LPPKALPTA---GLPGPLALGAEVARPPLAATHDIVOLVARALGSAQV 310
 Qy 329 ETNTMVP-----GSKFEELKHRRAARAAEAAHGTGPLAWDGAGFTSEDGRGG 381
 Db 311 QPKRLLPAPVNCQDLOQPGPESPGFLARFLANTSFOQRTGPV-WVTGSS PDEDSQ- 366
 Qy 382 ITLRVAVANGLGNNAKLLTRMQAGEAKLDFVEIMCPAG-CVGGGQOPRSKDATQKRO 439
 Db 367 -----CPAGOLCL-----DPGTNDSATLDAL 387
 Qy 440 QALYNYLDERKSTLR-----SHENPSIRELY-----DTVIGE 471
 Db 388 FAALANGSAPRALRKCCIGYCICDILLERLAEDTPPDEFLYLVGDKYGAIRDGRWTGLVD 447
 Qy 472 PLGKKAH 478
 Db 448 LLAGRAH 454

RESULT 3
 US-09-737-149-6
 ; Sequence 6, Application US/09737149
 ; Patent No. US2003007466A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Spaderna, Steven K
 ; APPLICANT: Quinn, Kerry E.
 ; APPLICANT: Shinkets, Richard A.
 ; APPLICANT: Muralidhara, Padigaru
 ; APPLICANT: Spytek, Kimberly A.
 ; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 15966-620 CIP
 ; CURRENT APPLICATION NUMBER: US/09/737,149
 ; CURRENT FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: 60/170,564
 ; PRIOR FILING DATE: 1999-12-14
 ; PRIOR APPLICATION NUMBER: 60/173,165
 ; PRIOR FILING DATE: 1999-12-27
 ; PRIOR APPLICATION NUMBER: 60/173,362
 ; PRIOR FILING DATE: 1999-12-29
 ; PRIOR APPLICATION NUMBER: 60/174,404
 ; PRIOR FILING DATE: 2000-01-04
 ; PRIOR APPLICATION NUMBER: 60/174,962
 ; PRIOR FILING DATE: 2000-01-07
 ; PRIOR APPLICATION NUMBER: 60/223,929
 ; PRIOR FILING DATE: 2000-08-09
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 6
 ; LENGTH: 971
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-737-149-6

Query Match 4.7%; Score 122; DB 10; Length 971;
 Best Local Similarity 21.9%; Pred. No. 0.1; Mismatches 120; Conservative 46; Indels 189; Gaps 22; Matches 120; Organism 46; Mismatches 189; Indels 192; Gaps 22;

RESULT 6
 US 09-934-070-6
 ; Sequence 6, Application US/09934070
 ; Publication No. US20030092004A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lipton, Stuart A.
 ; APPLICANT: Zhang, Dongxian
 ; APPLICANT: Charterton, Jon E.
 ; APPLICANT: Awobajo, Marc
 ; APPLICANT: Sevarino, Kevin A.
 ; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
 ; FILE REFERENCE: P-LJ 4900
 ; CURRENT APPLICATION NUMBER: US/09/934,070
 ; CURRENT FILING DATE: 2001-08-20
 ; NUMBER OF SEQ ID NOS: 53
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 1011
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909,
 ; LOCATION: 910, 931, 932, 933, 934, 935, 936, 937, 938, 943, 944, 945,
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 ; US-09-934-070-6

Query Match 4 6%; Score 118; DB 9; Length 1011;
 Best Local Similarity 21.8%; Pred. No. 0.24; Mismatches 128; Conservative 49; Mismatches 207; Indels 202; Gaps 25;

Qy 7 KPCAAVSIRSSCRARQAVAPRPLAASRVALTLEAPARRLGNVA-CAAAAPAAEP 64
 Db 25 QPGVGLARLGGSVRIGALLPAPLARARAAALARPLPILNLSLELWVAPRDP 84
 Qy 65 LSHVQ--QAL---AELAKPKKDP--TRKHYCVQVAPAVVIAETPLGLP-GATTP 112
 Db 85 ASLTRGICQALVPPGVAAALLAFAPEARPELQLHFLAAETPVLVSLRRBAPLPGAPNP 144
 Qy 113 KQLAGLRLRGDFEVDTLEGADLTIMEGSELLHRITLEAPRHSDFRPLPMTSCCG 172

RESULT 6
 US 09-934-070-6
 ; Sequence 6, Application US/09934070
 ; Publication No. US20030092004A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lipton, Stuart A.
 ; APPLICANT: Zhang, Dongxian
 ; APPLICANT: Charterton, Jon E.
 ; APPLICANT: Awobajo, Marc
 ; APPLICANT: Sevarino, Kevin A.
 ; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
 ; FILE REFERENCE: P-LJ 4900
 ; CURRENT APPLICATION NUMBER: US/09/934,070
 ; CURRENT FILING DATE: 2001-08-20
 ; NUMBER OF SEQ ID NOS: 53
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 1011
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909,
 ; LOCATION: 910, 931, 932, 933, 934, 935, 936, 937, 938, 943, 944, 945,
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 ; US-09-934-070-6

Query Match 4 6%; Score 118; DB 9; Length 1011;
 Best Local Similarity 21.8%; Pred. No. 0.24; Mismatches 128; Conservative 49; Mismatches 207; Indels 202; Gaps 25;

Qy 7 KPCAAVSIRSSCRARQAVAPRPLAASRVALTLEAPARRLGNVA-CAAAAPAAEP 64
 Db 25 QPGVGLARLGGSVRIGALLPAPLARARAAALARPLPILNLSLELWVAPRDP 84
 Qy 65 LSHVQ--QAL---AELAKPKKDP--TRKHYCVQVAPAVVIAETPLGLP-GATTP 112
 Db 85 ASLTRGICQALVPPGVAAALLAFAPEARPELQLHFLAAETPVLVSLRRBAPLPGAPNP 144
 Qy 113 KQLAGLRLRGDFEVDTLEGADLTIMEGSELLHRITLEAPRHSDFRPLPMTSCCG 172

RESULT 6
 US 09-934-070-6
 ; Sequence 6, Application US/09934070
 ; Publication No. US20030092004A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lipton, Stuart A.
 ; APPLICANT: Zhang, Dongxian
 ; APPLICANT: Charterton, Jon E.
 ; APPLICANT: Awobajo, Marc
 ; APPLICANT: Sevarino, Kevin A.
 ; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
 ; FILE REFERENCE: P-LJ 4900
 ; CURRENT APPLICATION NUMBER: US/09/934,070
 ; CURRENT FILING DATE: 2001-08-20
 ; NUMBER OF SEQ ID NOS: 53
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 1011
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909,
 ; LOCATION: 910, 931, 932, 933, 934, 935, 936, 937, 938, 943, 944, 945,
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 ; US-09-934-070-6

Query Match 4 6%; Score 118; DB 9; Length 1011;
 Best Local Similarity 21.8%; Pred. No. 0.24; Mismatches 128; Conservative 49; Mismatches 207; Indels 202; Gaps 25;

Qy 7 KPCAAVSIRSSCRARQAVAPRPLAASRVALTLEAPARRLGNVA-CAAAAPAAEP 64
 Db 25 QPGVGLARLGGSVRIGALLPAPLARARAAALARPLPILNLSLELWVAPRDP 84
 Qy 65 LSHVQ--QAL---AELAKPKKDP--TRKHYCVQVAPAVVIAETPLGLP-GATTP 112
 Db 85 ASLTRGICQALVPPGVAAALLAFAPEARPELQLHFLAAETPVLVSLRRBAPLPGAPNP 144
 Qy 113 KQLAGLRLRGDFEVDTLEGADLTIMEGSELLHRITLEAPRHSDFRPLPMTSCCG 172

RESULT 7
 US-09-937-149-3
 ; Sequence 33, Application US/09737149
 ; Patent No. US20020077466A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Spaderna, Steven K
 ; APPLICANT: Quinn, Kerry E.
 ; APPLICANT: Shinkets, Richard A.
 ; APPLICANT: Muralidhara, Padigaru
 ; APPLICANT: Sypert, Kimberly A.
 ; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 15966-620 CIP
 ; CURRENT APPLICATION NUMBER: US/09/737,149
 ; CURRENT FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: 60/170,564
 ; PRIOR FILING DATE: 1999-12-14
 ; PRIOR APPLICATION NUMBER: 60/173,165
 ; PRIOR FILING DATE: 1999-12-27
 ; PRIOR APPLICATION NUMBER: 60/173,362
 ; PRIOR FILING DATE: 1999-12-27
 ; PRIOR APPLICATION NUMBER: 60/173,544
 ; PRIOR FILING DATE: 1999-12-29
 ; PRIOR APPLICATION NUMBER: 60/174,404
 ; PRIOR FILING DATE: 2000-01-04
 ; PRIOR APPLICATION NUMBER: 60/174,962
 ; PRIOR FILING DATE: 2000-01-07
 ; PRIOR APPLICATION NUMBER: 60/223,929
 ; PRIOR FILING DATE: 2000-08-09
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 33
 ; LENGTH: 901
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-937-149-3

Query Match 4 5%; Score 116; DB 10; Length 901;
 Best Local Similarity 21.7%; Pred. No. 0.31; Mismatches 128; Conservative 49; Mismatches 207; Indels 206; Gaps 25;

Qy 7 KPCAAVSIRSSCRARQAVAPRPLAASRVALTLEAPARRLGNVA-CAAAAPAAEP 64
 Db 25 QPGVGLARLGGSVRIGALLPAPLARARAAALARPLPILNLSLELWVAPRDP 84

US-09-737-149-8

Query Match 4 5%; Score 116; DB 10; Length 1043;
Best Local Similarity 21 7%; Pred. No. 0 38;
Matches 128; Conservative 49; Mismatches 207; Indels 206; Gaps 25;

QY 65 LSHVQ---QAL-----AELAKPKDP--TRKHCVQVAPAVRVAIAETGLAP-GATTP 112
Db 85 ASLTRLGLCQALVPGVAAALLAFAEARPELQLQHHTAATETPVLSLREARALGAPN 144

QY 113 KOLAEGLRLGFDEVDTLFAGDLTMEGSELLHLTEHEAHPHSDEPLPMTSCCPG 172
Db 145 FHL---OLHWASPLET-----LLDVLVAVLQAHAWEDVGLALCRRQDPG 185

QY 173 WIAMLERSKY---PDLIPYVSSCKSPQMLLAAMVSKYLAEKKGIAPKDMVMSIMPCTRK 228
Db 186 GLVALWTSRAGRPPQVLVIDLSRDRDGAGLRAPIALPMAAPVGGAP----- 231

QY 229 QSEADRDWFCVDADPTLRLQDHVITVLELG-NIFKERGINLAELPEGMNDNPMSGAG 287
Db 232 -----VPAAVLGLGCDIARRVLEAVPPGPHW----- 258

QY 288 VLFGTTGGVMEAALRTAYELFTGTPPLPLRSLSEVR-----GMDGIK 328
Db 259 -LIGTP--LPPKALPTA---GLPPGLLALGEVARPPLAATHDIVOLVARALGSAQV 310

QY 329 ETNTITAVPAP-----GSKEELKHRAA 351
Db 311 QPKRALLPAPVNGCQLQPGPESPGRLARFLANISFQGRTGPVWVTSQVHMSRHKV 370

QY 352 ARAEEAAAHGTPGLPLANDGGAGFTSEDR---GGITLRLVAVANGLGNNAKL--ITKMQA 404
Db 371 WSLRDRPGRGAP--AWATVGSW--RDGQDLDEPGGASARPPPGQVWPKLRYVTLLE- 424

QY 405 GEAKYDFV-----ETMACPAG--CVGGGGQPRSTDKAITOKRQALYNLDEKSTLR-- 454
Db 425 ---HFPVFAARDPDEDGQCPAGQCLL---DPGTNDSATLDAFLAFAALANGSAPRALRKCC 476

QY 455 -----SHENPSIREY-----DTYLGEPGKHAK 478
Db 477 YGYCIDDLERLAEDTPFDDELYLVGDGYKAGLDRGRTWGLVGDLLAGRAH 526

RESULT 8
US-09-377-149-B

Patent No. US/09/737149

GENERAL INFORMATION:

APPLICANT: Spadeine, Steven K
APPLICANT: Quinn, Kerry E.
APPLICANT: Shinkets, Richard A.
APPLICANT: Muralidhara, Radigaru
APPLICANT: Spytek, Kimberly A.

TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
FILE REFERENCE: 15966620 CIP
CURRENT APPLICATION NUMBER: US/09/737,149

CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/170,564
PRIOR FILING DATE: 1999-12-14

PRIOR APPLICATION NUMBER: 60/173,165
PRIOR FILING DATE: 1999-12-27
PRIOR APPLICATION NUMBER: 60/173,362
PRIOR FILING DATE: 1999-12-27
PRIOR APPLICATION NUMBER: 60/173,544
PRIOR FILING DATE: 1999-12-29

PRIOR APPLICATION NUMBER: 60/174,404
PRIOR FILING DATE: 2000-01-04
PRIOR APPLICATION NUMBER: 60/174,962
PRIOR FILING DATE: 2000-01-07
PRIOR APPLICATION NUMBER: 60/223,929
PRIOR FILING DATE: 2000-08-09

NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1043

TYPE: PRT
ORGANISM: Homo sapiens

US-09-737-149-8

Query Match 4 5%; Score 116; DB 10; Length 1043;
Best Local Similarity 21 7%; Pred. No. 0 38;
Matches 128; Conservative 49; Mismatches 207; Indels 206; Gaps 25;

QY 7 KPCAavisrisscraQVaPAPLAAStVrVaLATEAPARRIGNA-CAAAPAPAEAP 64
Db 25 QPGVGLARLGGSVRLGALLPRAPIALARALARALAPRLPHNLSLELVWAAPPDRP 84

QY 113 KOLAEGLRLGFDEVDTLFAGDLTMEGSELLHLTEHEAHPHSDEPLPMTSCCPG 172
Db 145 FHL---OLHWASPLET-----LLDVLVAVLQAHAWEDVGLALCRRQDPG 185

QY 173 WIAMLERSKY---PDLIPYVSSCKSPQMLLAAMVSKYLAEKKGIAPKDMVMSIMPCTRK 228
Db 186 GLVALWTSRAGRPPQVLVIDLSRDRDGAGLRAPIALPMAAPVGGAP----- 231

QY 229 QSEADRDWFCVDADPTLRLQDHVITVLELG-NIFKERGINLAELPEGMNDNPMSGAG 287
Db 232 -----VPAAVLGLGCDIARRVLEAVPPGPHW----- 258

QY 288 VLFGTTGGVMEAALRTAYELFTGTPPLPLRSLSEVR-----GMDGIK 328
Db 259 -LIGTP--LPPKALPTA---GLPPGLLALGEVARPPLAATHDIVOLVARALGSAQV 310

QY 329 ETNTITAVPAP-----GSKEELKHRAA 351
Db 311 QPKRALLPAPVNGCQLQPGPESPGRLARFLANISFQGRTGPVWVTSQVHMSRHKV 370

QY 352 ARAEEAAAHGTPGLPLANDGGAGFTSEDR---GGITLRLVAVANGLGNNAKL--ITKMQA 404
Db 371 WSLRDRPGRGAP--AWATVGSW--RDGQDLDEPGGASARPPPGQVWPKLRYVTLLE- 424

QY 405 GEAKYDFV-----ETMACPAG--CVGGGGQPRSTDKAITOKRQALYNLDEKSTLR-- 454
Db 425 ---HFPVFAARDPDEDGQCPAGQCLL---DGTNDSATLDAFLAFAALANGSAPRALRKCC 476

QY 455 -----SHENPSIREY-----DTYLGEPGKHAK 478
Db 477 YGYCIDDLERLAEDTPFDDELYLVGDGYKAGLDRGRTWGLVGDLLAGRAH 526

RESULT 9
US-09-976-059-15

Sequence 15, Application US/09976059
Patent No. US/02020164747A1

GENERAL INFORMATION:

APPLICANT: Farnet, Chris
APPLICANT: Zazopoulos, Emmanuel
APPLICANT: Staffa, Alfredo

TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
FILE REFERENCE: 3019-PCT
CURRENT APPLICATION NUMBER: US/09/976,059
CURRENT FILING DATE: 2001-10-15
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.0
SEQ ID NO 15
LENGTH: 4999

TYPE: PRT
ORGANISM: Actinoplanes sp.

US-09-976-059-15

Query Match 4 5%; Score 116; DB 9; Length 4999;
Best Local Similarity 24 18%; Pred. No. 3 5; Mismatches 171; Indels 112; Gaps 20;
Matches 100; Conservative 32; Mismatches 171; Indels 112; Gaps 20;

QY 8 PQAavstrgsscrarQVa--PAPLAAStVrVaLATEAPARRLGWVACAAAPAEAP 64

Db 371 VAAATLSIRYIA-DROQPKAIDLIDBAASSRMOIDSKEBLDRDRRIOLKQQA 428 Qy 343 EEL------LKHRAARAEEA-----AHGTPGPA-----W 367
 Qy 200 AMVSKYLAEEKKGIAPKDMVMSIMPCTRKQOSEADRWDWCVDAD---PTLR-OLDHVVIT 254 Db 298 VELJARALSSMAJNHPERALLPAAVNCEDIKTGGESPARTLARWELTSFQRTGAVW 357
 Db 429 LMKESDEASKRLL---DMNEELSDKERQYSELEENWAKASLSGTOQIA 485 Qy 368 DGGAGFTS-----EDGR----GGITLRVAVANGLGNNA 396
 Qy 255 VE---LGNIFKERGINLAELPEGEWDMPMGVSGAGVLFGT---GGVMAALRFA 305 Db 358 VAGSSQVHVSRLRDKPLGAPAWTVGSKQDGDFQPGAAALRVPSPSGTOQRP 417
 Db 486 IEGARRVGDALARSELQGKIPSEKOL---EAATOLEGKHMRLRNKVDAELAVL 540 Qy 397 KL-ITKMOAGEAKYDFV---EIMACPAG-CVGGGQPRSTDRAITKQQAALYN 446
 Qy 306 ELEGTGPPLSLSEVRHDGKETNITMPAPGSKFELLKHRAAEEA----- 356 Db 418 KLRYVTLVE---HPFVFTRESDEDGOCPAGQCL---DPCNDSARLDALFTALEG 468
 Db 541 ARMTGIPPSRMMESE-----REKLRM-----EQELHHRVIGQNEADAVSNAR 585 Qy 447 DEKSLRR 454
 Qy 357 -AAHGTPPLAWPGAGFTSESDRGGITLTVANGLGNNAKLTQMAGE- AKYDVE 413 Db 469 SVPTLRR 476
 Db 586 RSRAGLADPNRPGSFLFLGPGVKGTECKALANFMFDSDDEAMVRIDMSEFMKHSVSR 645

RESULT 12 US-10-188-246-12
 ; Sequence 8, Application US/09334070 ; Sequence 12, Application US/10188246
 ; Publication No. US20030087274A1 ; Publication No. US20030087274A1
 ; GENERAL INFORMATION: ; GENERAL INFORMATION:
 ; APPLICANT: Lipton, Stuart A. ; APPLICANT: Anderson, David W.,
 ; APPLICANT: Zhang, Dongqian ; APPLICANT: Bodog, Ferenc,
 ; APPLICANT: Chatteerton, Jon E. ; APPLICANT: Casman, Stacie,
 ; APPLICANT: Avobullyi, Marc ; APPLICANT: Edinger, Shlomit,
 ; APPLICANT: Sevarino, Kevin A. ; APPLICANT: Gerlach, Valerie,
 ; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS ; APPLICANT: Goman, Linda,
 ; FILE REFERENCE: P-WJ 4900 ; APPLICANT: Li, Li,
 ; CURRENT APPLICATION NUMBER: US/09/934,070 ; APPLICANT: Malivankar, Uriel,
 ; NUMBER OF SEQ ID NOS: 53 ; APPLICANT: Patturajan, Meera,
 ; SOFTWARE: FASTSEQ for Windows Version 4.0 ; APPLICANT: Peiman, John,
 ; SEQ ID NO 8 ; APPLICANT: Shenoy, Suresh,
 ; LENGTH: 1005 ; APPLICANT: Shinkets, Richard,
 ; TYPE: PRT ; APPLICANT: Vernet, Corinne A. M.,
 ; ORGANISM: Mus musculus ; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND MET
 ; US-09-934-070-8 ; FILE REFERENCE: 21402-39B US
 ; PRIORITY APPLICATION NUMBER: 60/303046 ; CURRENT APPLICATION NUMBER: US/10/188,246
 ; PRIORITY FILING DATE: 2002-07-02 ; CURRENT FILING DATE: 2002-07-02
 ; PRIORITY FILING DATE: 2001-07-05 ; PRIORITY APPLICATION NUMBER: 60/303046
 ; PRIORITY FILING DATE: 2001-07-11 ; PRIORITY FILING DATE: 2001-07-11
 ; PRIORITY FILING DATE: 2001-07-12 ; PRIORITY FILING DATE: 2001-07-12
 ; PRIORITY FILING DATE: 2001-07-17 ; PRIORITY FILING DATE: 2001-07-17
 ; PRIORITY APPLICATION NUMBER: 60/326981 ; PRIORITY APPLICATION NUMBER: 60/326981
 ; PRIORITY FILING DATE: 2001-07-14 ; PRIORITY FILING DATE: 2001-07-14
 ; PRIORITY APPLICATION NUMBER: 60/360923 ; PRIORITY APPLICATION NUMBER: 60/360923
 ; PRIORITY FILING DATE: 2002-02-28 ; PRIORITY FILING DATE: 2002-02-28
 ; PRIORITY APPLICATION NUMBER: 60/363636 ; PRIORITY APPLICATION NUMBER: 60/363636
 ; PRIORITY FILING DATE: 2002-03-12 ; PRIORITY FILING DATE: 2002-03-12
 ; PRIORITY APPLICATION NUMBER: 60/373063 ; PRIORITY FILING DATE: 2002-04-16
 ; PRIORITY FILING DATE: 2002-04-16 ; NUMBER OF SEQ ID NOS: 60
 ; SOFTWARE: Custom ; SEQ ID NO 12
 ; LENGTH: 450 ; LENGTH: 450
 ; TYPE: PRT ; TYPE: PRT
 ; ORGANISM: Homo sapiens ; ORGANISM: Homo sapiens
 ; US-10-188-246-12 ; US-10-188-246-12

Query Match 4.2%; Score 107.5; DB 9; Length 450; Qy 343 EEL------LKHRAARAEEA-----AHGTPGPA-----W 367
 Best Local Similarity 20.5%; Pred. No. 0.61; Indels 151; Gaps 20; Matches 92; Conservative 53; Mismatches 152; Indels 151; Gaps 20;

Query Match 4.2%; Score 107.5; DB 9; Length 450; Qy 343 EEL------LKHRAARAEEA-----AHGTPGPA-----W 367
 Best Local Similarity 20.5%; Pred. No. 0.61; Indels 151; Gaps 20; Matches 92; Conservative 53; Mismatches 152; Indels 151; Gaps 20;

Db 230 SEADRDWFCVDAADTILRQLDHVTTVEGLNIFKERGINLAELPEGEWDMPMGVSGAGVLT 289 Qy 17 SSCHARQVAPRPL-AASTRVVAL-TLAP-----ARRGNVACAAAPAAEAPL 65
 Db 197 SQARK--FVLD---LSQDS---GNDSLRATLALLTLEGG---GIPVSAVL 238 Qy 15 SAGR----PRDGIERAAVLRAGAGPVRSPGGGGGGGRTIAQAGAAVPAAVPR 69
 Qy 290 FGTGIGG---VMEALRRAVELEGTGPPLSLSEV---RGMGQIKETNITMPAPGK 342 Db 66 SHVQOALAEALKPKDPTRKHVCQV-----PAVRAVIA-----AETL-GLAPGAT 110
 Db 239 LGCTTAHNEHVELEAA-PGPQWQLGTPLPAAELPKTGPQGVLVIGTCOPSLAAVIDM 297

RESULT 14
US-09-934-070-2

Sequence 2: Application US/09934070
Publication No. US20030092004A1
GENERAL INFORMATION:
APPLICANT: Lipton, Stuart A.
APPLICANT: Zhang, Dongxian
APPLICANT: Chatterton, Jon E.
APPLICANT: Awobuluyi, Marc
APPLICANT: Sevarino, Kevin A.
TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
FILE REFERENCE: P-LJ 4900
CURRENT APPLICATION NUMBER: US/09/934,070
NUMBER OF SEQ ID NOS: 53
CURRENT FILING DATE: 2001-08-20
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 2
LENGTH: 987
TYPE: PRT
ORGANISM: Rattus sp.
US-09-934-070-2

Query Match 4.28; Score 107.5; DB 9; Length 987;
Best Local Similarity 22.0%; Pred. No. 1.9; Matches 114; Conservative 48; Mismatches 200; Indels 157; Gaps 23; Qy 7 KPCAAVSISSCRARQVAPRPLAATRVALTLEAPARRGVACAAAPAEAPLS 66
Db 27 QPCRVPTRAGASVRLAALPRAPARARVLAALAT--PAPRLPH----- 68
Qy 67 HVOQALAEALKPKDPT--RKHVQVQVAVAVRAIAETGLAGPATTPKQLAEGIIRRGLF 124
Db 69 NLSLELVAVASPTRDPSALRGICQVLP-----GVASIAFPARPE-LRLQF 118
Qy 123 ---DEVFTLFGADLTIMEGSEJLHRTELEAHPHSDEPLPMFTSCCPGWTAMLEKS 180
Db 119 LAAMETPVFPHLDWASPLETILDVLYSLVRAHAWDIALYLCRVDPGSLVLTWN 178
Qy 181 Y---PDLTIVYSSCKSPORNMLAAMVSKYLAEGKGIAPKDMVMSIMPC 236
Db 179 HASQAKPFVLDLSRDSRNSLRAGLALIGALEGGGTPVPAV--LLGOSTARAH----- 232
Qy 237 FCVADADPTLQDLDHIVTVELGNIFKERCINLAELPEGW--DNPM-----GVGSCA 286
Db 233 -VLEAP-----PGPMLLSTPLPAPALPPTGGLPPCV 263

RESULT 15
US-09-934-070-4

Sequence 4: Application US/09934070
Publication No. US20030092004A1
GENERAL INFORMATION:
APPLICANT: Lipton, Stuart A.
APPLICANT: Zhang, Dongxian
APPLICANT: Chatterton, Jon E.
APPLICANT: Awobuluyi, Marc
APPLICANT: Sevarino, Kevin A.
TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS
FILE REFERENCE: P-LJ 4900
CURRENT APPLICATION NUMBER: US/09/934,070
CURRENT FILING DATE: 2001-08-20
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 1002
TYPE: PRT
ORGANISM: Rattus sp.
US-09-934-070-4

Query Match 4.0%; Score 104; DB 9; Length 1002;
Best Local Similarity 22.1%; Pred. No. 3.8; Matches 117; Conservative 48; Mismatches 201; Indels 164; Gaps 25; Qy 7 KPCAAVSISSCRARQVAPRPLAATRVALTLEAPARRGVACAAAPAEAPLS 66
Db 27 QPCRVPTRAGASVRLAALPRAPARARVLAALAT--PAPRLPH----- 68
Qy 67 HVOQALAEALKPKDPT--RKHVQVQVAVAVRAIAETGLAGPATTPKQLAEGIIRRGLF 124
Db 69 NLSLELVAVASPTRDPSALRGICQVLP-----GVASIAFPARPE-LRLQF 118
Qy 115 LAEGIRR-----GFDDEVFTLFGADLTIMEGSEJLHRTELEAHPHSDEPLPMFTSC 169
Db 127 WSVSIRREVFTALG---APTPFHQLDWSPLTFLDVLVSLVRAHAMEDIALVLCRV 182
Qy 170 CPGWTAMLEKSY---PDLIPIVSSCKSPORNMLAAMVSKYLAEGKGIAPKDMVMSIMPC 225
Db 183 DPGSPLVLTWNHASPKPFLDSRDSRNSLRAGLALIGALEGGGTPVPAV--LLGIC 240
Qy 226 TRKSEADRDWFCDADPLTQDLDHIVTVELGNIFKERCINLAELPEGW--DNPM--- 280
Db 241 STARAH-----VLEAP-----PGPMLLSTPLPAPALPPTGGLPPCV 267
Qy 281 ---GVGSGAGVLEGGTGGVMEALRTAYELF----- 308
Db 268 ALPTTGLPPGVLALGETEOLSLEAVVHDAVELVQALSSMALVPERALLPAVNCDDLK 327
Qy 309 -----TGTPLPR-LSLSEVRGMGCKIENITMFPAGSKFEELKHKRAARAEEAAGT 361
Db 328 TGGSATGTRLARELNGTSRQGRG-----AWWTGSSQVHRSRHKWMSLRDPGLA 380
Qy 362 PGPLW-----DGGAGFTSEDGRRGGITLRAVAVANGLGNNAKKL--ITKMQAGEAKYDFV 412

wed Jun 4 16:29:41 2003

cac80065.rapb

Page 9

Db 381 P---AWATVGSMQDGQDFQ---| GAAALRVPSGQARPKLRLVVTIVE----: |||
QY 413 -----ELMACPAG-CVGGGGOPRSIDKAIQKQALYNLDEKSIRR 454
Db 429 FTRESDEPGCQCPAGLCI-----DPGTNSARLDALEAALVNGSYPRTRR 474

Search completed: June 3, 2003, 16:17:40
Job time : 26 secs

This Page Blank (uspto)

PR	15-DEC-1999;	99US-0172367.
PA	(INCY-)	INCYTE GENOMICS INC.
PI	Yue H, Lal P, Tang YT, Hillman JL, Baughn MR, Azimzai Y, Lu DAM;	
XX	WPI: 2001-390245/41.	
DR	N-PSDB; AAH24245.	
XX	Novel human oxidoreductase protein (ORP) useful for diagnosing, treating and preventing cell proliferative, neurological, viral, reproductive and autoimmune/inflammatory disorders associated with abnormal expression of ORP	
PT		
PT	Novel human oxidoreductase protein (ORP) useful for diagnosing, treating and preventing cell proliferative, neurological, viral, reproductive and autoimmune/inflammatory disorders associated with abnormal expression of ORP	
PT		
XX	Claim 1; Page 116-117; 136pp; English.	
PS		
XX	Sequences AAB73668-AAB73694 represent 27 novel human oxidoreductase proteins, designated ORP-1 to ORP-27 respectively, and sequences AAB4223-AAH24249 represent cDNAs encoding ORP-1 to ORP-27. Human ORP proteins and nucleic acids are useful for diagnosing, treating or preventing cell proliferative disorders (e.g. arteriosclerosis, cirrhosis, psoriasis, cancers); endocrine disorders (e.g., type I or II diabetes mellitus, diabetes insipidus, dwarfism, hirsutism, amenorrhea, osteoporosis); metabolic disorders (e.g., obesity, phenylketonuria, hypercholesterolemia); reproductive disorders (e.g., infertility, ovulatory and menstrual cycle defects, endometriosis, polycystic ovary disease, disruption of spermatogenesis, impotence); neurological disorders (e.g., epilepsy, stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease, meningitis, Creutzfeldt-Jakob disease, cerebral palsy, muscular dystrophy, mood, anxiety and schizophrenic disorders); viral, bacterial, fungal and parasitic infections; and autoimmune/inflammatory disorders such as acquired immunodeficiency syndrome (AIDS), allergies, asthma, Crohn's disease, atopic dermatitis, gout, multiple sclerosis, rheumatoid arthritis or ulcerative colitis. Human ORP proteins and nucleotides can be used to identify compounds which modulate their activity or expression. ORP nucleic acid sequences may also be used for assessing the toxicity of a test compound, to detect upstream sequences such as promoters and regulatory elements, and to create knock out or knock in animals or transgenic animals to model human disease. Oligonucleotide primers derived from ORP gene sequences may be used to detect single nucleotide polymorphisms (SNPs) and for mapping the naturally occurring genomic sequences. Antibodies specific for ORP proteins may be used in the diagnosis of disorders associated with aberrant ORP expression, in assays to monitor patients being treated with ORP or modulators thereof, and for assessing toxicity of potential CC drugs.	
XX	Sequence 476 AA;	
SO		
Query Match	17.7%; Score 456.5; DB 22; Length 476;	
Best Local Similarity	27.8%; Pred. No. 1.5e-36;	
Matches	126; Conservative 82; Mismatches 168; Indels 77; Gaps 15;	
Qy	52 VACAAAPAAEAPLSPHVVQALAEIAKPKD----DPTRKH-VCVQVAPAVRAIAETLGL 105	
Qy	106 APTATPKQALAEGLRGLFDEVFTLFGADLTIMPEGESELLHLRTEAHPHSDEPPLM 165	
Qy	131 NP-TTARKITSFFRKIGVHFVDFAFSRSLSESQEVRRFGQACR---QALPL 185	
Db	166 FTSCCGPGWIMLAKMVKYPD-LIPYVSSCKSPOMMLAMVSKYLAERKGIAKDMVNVIMP 224	
Db	186 LASACPGWICVAKTHGSFLPHSTARSPOQVMGLSVKQHLPDKIVHTVMP 245	
Db	225 CTRKSEADR-DWFCVWADPLTLRQDHDVITVLEGLNIFKERGINIAELPGEWDDNPMGV 283	
Db	246 CYDKLEASRPDF--NQEQHOTRDVCVLTGGEVRLLEHEGVSPDPLRPLS--LC 300	
Db	284 SGAGY-----LFGTGGVVAEALR-TAYLFLFGTPLPRLSVEVRGMDCIKETNTMVF 336	
Db	301 SGASAEPEETSHRGGSGGYLEHVFHRARLE-----GIVHVAETVYK 343	
XX	Query Match	14.0%; Score 361.5; DB 22; Length 456;
XX	Best Local Similarity	24.3%; Pred. No. 5e-27;
XX	Matches	112; Conservative 76; Mismatches 174; Indels 99; Gaps 13;
Qy	52 VACAAAPAAEAPLSPHVVQALAEIAKPKD----DPTRKH-VCVQVAPAVRAIAETLGL 106	
Qy	54 LACDSCTAEGVYQLSQONAKDFRVRVNLKCKDTSKHKVLUVSVCPOLSPYFAAKNL 113	
Db	107 PGATTPKQALAEGLRGLFDEVFTLFGADLTIMPEGESELLHLRTEAHPHSDEPPLM 166	
Db	114 V-TDASRLLCGFLKLSLVHVVYFDTTIAADFSLSESQEVRRYQHSE---EERLPLM 168	

QY	167	TSCCPGMIALEKSY PDLIPVSSCKSPOMMLAAMWKSYLAEKKGAKPDAMVNSIMPC	225
Db	169	TSACPGWVRYAERVLGRPITHLCTAKSPQOMGLVSKYDIFARQONLSPEKIFHVTIAPC	228
CC	226	TRKQSEADRDNFCVADPTL--RQDHVITVVELGNIFKERGINLAELPEGEWDNPNGV	282
CC	227	TRKQSEADRDNFCVADPTL--RQDHVITVVELGNIFKERGINLAELPEGEWDNPNGV	282
CC	229	YDKKLEALQE---SIPPALHOSRGACVLTSGEIAQIMEQDLSVRD-----	272
Db	283	GSGAGVLFG-----TTGGMEEALRTAVELFTGTPPLPRLSLEVRGMDCIK	328
QY	346	ANGLWNAKKLTQMAGEAKTDFVEIMACPACGCVGGGQPRS---TDKAITQRQALY	444
Db	406	AYGFRNIQNMILKKGKFPFHFVVELACAGCCCLNGRQAQTPDGHADKALIQRomegiya	405
QY	445	NLDEKSTLRRSHENPSPRELVTYLGPLGHKAELHHTY	485
Db	406	DI---PVRRPESAAHVOELQEWLEGINSPKAREVLHTY	442
RESULT 3			
ID	AAB92498	standard; Protein; 456 AA.	
XX			
AC	AAB92498;		
XX			
DT	26-JUN-2001	(first entry)	
XX			
DE	Human protein sequence	SEQ ID NO:10502.	
XX			
OS	Homo sapiens.		
XX			
PN	EP1074617-A2.		
XX			
PD	07-FEB-2001.		
XX			
PF	28-JUL-2000;	2000EP-0116126.	
XX			
PR	29-JUL-1999;	99JP-0248036.	
PR	27-AUG-1999;	99JP-0300053.	
PR	11-JAN-2000;	2000JP-0118776.	
PR	02-MAY-2000;	2000JP-01183767.	
PR	09-JUN-2000;	2000JP-0241899.	
PA	(HELI-)	HELIX RES INST.	
XX			
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX			
DR	WPI; 2001-318749/34.		
XX			
PS	Claim 8; SEQ ID 10602; 2537pp + CD ROM; English.		
PT	primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -		
PT	The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end		
RESULT 4			
ID	ABB69383	standard; Protein; 241 AA.	
XX			
AC	ABB69383;		
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE	Drosophila melanogaster polypeptide	SEQ ID NO 34941.	
XX			
KW	Drosophila; developmental biology; cell signalling; insecticide;		
XX			
OS	Drosophila melanogaster.		
PN	WO20011042-A2.		
XX			
PD	27-SEP-2001.		
XX			
Sequence	456 AA;		
Query	PGATTPKOLAGLRRGFDEFTLFGADLTMEGSELLRLTLEAHPRHDEPLPMPF	14.0%; Score 361.5; DB 22; Length 456;	
Matches	112; Conservative	24.3%; Pred. No. 27; Mismatches 174; Indels 99; Gaps 13;	
Db	114	V-TDASRRLCGFLKSLGVHVVFTTAADESFILESDRFTTAAESRERVRYRHSE--EERTLPLM	168
QY	52	VACAAAPAPATAPLSHVQQA-----LAELAKPKDDPRTKRVHQVAVAPVRAIAETGLA	106
Db	54	LACDSCTAAEGVQVLSQONAKDFRVLNLNKCDTSKHKVVLVSVCPQSLVYAAKFLS	113
QY	107	TRKQSEADRDNFCVADPTL--RQDHVITVVELGNIFKERGINLAELPEGEWDNPNGV	225
Db	169	TSACPGWVRYAERVLGRPITHLCTAKSPQOMGLAAMWKSYLAEKKGAKPDAMVNSIMPC	228
QY	226	YDKKLEALQE---SIPPALHOSRGACVLTSGEIAQIMEQDLSVRD-----	272
Db	283	GSGAGVLFG-----TTGGMEEALRTAVELFTGTPPLPRLSLEVRGMDCIK	328
QY	346	ANGLWNAKKLTQMAGEAKTDFVEIMACPACGCVGGGQPRS---TDKAITQRQALY	444
Db	406	AYGFRNIQNMILKKGKFPFHFVVELACAGCCCLNGRQAQTPDGHADKALIQRomegiya	405
QY	445	NLDEKSTLRRSHENPSPRELVTYLGPLGHKAELHHTY	485
Db	406	DI---PVRRPESAAHVOELQEWLEGINSPKAREVLHTY	442
Sequence	456 AA;		
Query	Match	Score	
Best	Local	Similarity	
Matches	112;	Conservative	
Db	114	Mismatches	
QY	52	Indels	
Db	54	Gaps	
QY	107	Length	
Db	114		
QY	167		
Db	169		
QY	226		
Db	229		
QY	283		
Db	273		
QY	329		
Db	317		
QY	389		
Db	346		
QY	445		
Db	406		
Sequence	456 AA;		
Query	PGATTPKOLAGLRRGFDEFTLFGADLTMEGSELLRLTLEAHPRHDEPLPMPF	14.0%; Score 361.5; DB 22; Length 456;	
Matches	112; Conservative	24.3%; Pred. No. 27; Mismatches 174; Indels 99; Gaps 13;	
Db	114	V-TDASRRLCGFLKSLGVHVVFTTAADESFILESDRFTTAAESRERVRYRHSE--EERTLPLM	168
QY	52	VACAAAPAPATAPLSHVQQA-----LAELAKPKDDPRTKRVHQVAVAPVRAIAETGLA	106
Db	54	LACDSCTAAEGVQVLSQONAKDFRVLNLNKCDTSKHKVVLVSVCPQSLVYAAKFLS	113
QY	107	TRKQSEADRDNFCVADPTL--RQDHVITVVELGNIFKERGINLAELPEGEWDNPNGV	225
Db	169	TSACPGWVRYAERVLGRPITHLCTAKSPQOMGLAAMWKSYLAEKKGAKPDAMVNSIMPC	228
QY	226	YDKKLEALQE---SIPPALHOSRGACVLTSGEIAQIMEQDLSVRD-----	272
Db	283	GSGAGVLFG-----TTGGMEEALRTAVELFTGTPPLPRLSLEVRGMDCIK	328
QY	346	ANGLWNAKKLTQMAGEAKTDFVEIMACPACGCVGGGQPRS---TDKAITQRQALY	444
Db	406	AYGFRNIQNMILKKGKFPFHFVVELACAGCCCLNGRQAQTPDGHADKALIQRomegiya	405
QY	445	NLDEKSTLRRSHENPSPRELVTYLGPLGHKAELHHTY	485
Db	406	DI---PVRRPESAAHVOELQEWLEGINSPKAREVLHTY	442
Sequence	456 AA;		
Query	Match	Score	
Best	Local	Similarity	
Matches	112;	Conservative	
Db	114	Mismatches	
QY	52	Indels	
Db	54	Gaps	
QY	107	Length	
Db	114		
QY	167		
Db	169		
QY	226		
Db	229		
QY	283		
Db	273		
QY	329		
Db	317		
QY	389		
Db	346		
QY	445		
Db	406		
Sequence	456 AA;		
Query	PGATTPKOLAGLRRGFDEFTLFGADLTMEGSELLRLTLEAHPRHDEPLPMPF	14.0%; Score 361.5; DB 22; Length 456;	
Matches	112; Conservative	24.3%; Pred. No. 27; Mismatches 174; Indels 99; Gaps 13;	
Db	114	V-TDASRRLCGFLKSLGVHVVFTTAADESFILESDRFTTAAESRERVRYRHSE--EERTLPLM	168
QY	52	VACAAAPAPATAPLSHVQQA-----LAELAKPKDDPRTKRVHQVAVAPVRAIAETGLA	106
Db	54	LACDSCTAAEGVQVLSQONAKDFRVLNLNKCDTSKHKVVLVSVCPQSLVYAAKFLS	113
QY	107	TRKQSEADRDNFCVADPTL--RQDHVITVVELGNIFKERGINLAELPEGEWDNPNGV	225
Db	169	TSACPGWVRYAERVLGRPITHLCTAKSPQOMGLAAMWKSYLAEKKGAKPDAMVNSIMPC	228
QY	226	YDKKLEALQE---SIPPALHOSRGACVLTSGEIAQIMEQDLSVRD-----	272
Db	283	GSGAGVLFG-----TTGGMEEALRTAVELFTGTPPLPRLSLEVRGMDCIK	328
QY	346	ANGLWNAKKLTQMAGEAKTDFVEIMACPACGCVGGGQPRS---TDKAITQRQALY	444
Db	406	AYGFRNIQNMILKKGKFPFHFVVELACAGCCCLNGRQAQTPDGHADKALIQRomegiya	405
QY	445	NLDEKSTLRRSHENPSPRELVTYLGPLGHKAELHHTY	485
Db	406	DI---PVRRPESAAHVOELQEWLEGINSPKAREVLHTY	442
Sequence	456 AA;		
Query	Match	Score	
Best	Local	Similarity	
Matches	112;	Conservative	
Db	114	Mismatches	
QY	52	Indels	
Db	54	Gaps	
QY	107	Length	
Db	114		
QY	167		
Db	169		
QY	226		
Db	229		
QY	283		
Db	273		
QY	329		
Db	317		
QY	389		
Db	346		
QY	445		
Db	406		
Sequence	456 AA;		
Query	PGATTPKOLAGLRRGFDEFTLFGADLTMEGSELLRLTLEAHPRHDEPLPMPF	14.0%; Score 361.5; DB 22; Length 456;	
Matches	112; Conservative	24.3%; Pred. No. 27; Mismatches 174; Indels 99; Gaps 13;	
Db	114	V-TDASRRLCGFLKSLGVHVVFTTAADESFILESDRFTTAAESRERVRYRHSE--EERTLPLM	168
QY	52	VACAAAPAPATAPLSHVQQA-----LAELAKPKDDPRTKRVHQVAVAPVRAIAETGLA	106
Db	54	LACDSCTAAEGVQVLSQONAKDFRVLNLNKCDTSKHKVVLVSVCPQSLVYAAKFLS	113
QY	107	TRKQSEADRDNFCVADPTL--RQDHVITVVELGNIFKERGINLAELPEGEWDNPNGV	225
Db	169	TSACPGWVRYAERVLGRPITHLCTAKSPQOMGLAAMWKSYLAEKKGAKPDAMVNSIMPC	228
QY	226	YDKKLEALQE---SIPPALHOSRGACVLTSGEIAQIMEQDLSVRD-----	272
Db	283	GSGAGVLFG-----TTGGMEEALRTAVELFTGTPPLPRLSLEVRGMDCIK	328
QY	346	ANGLWNAKKLTQMAGEAKTDFVEIMACPACGCVGGGQPRS---TDKAITQRQALY	444
Db	406	AYGFRNIQNMILKKGKFPFHFVVELACAGCCCLNGRQAQTPDGHADKALIQRomegiya	405
QY	445	NLDEKSTLRRSHENPSPRELVTYLGPLGHKAELHHTY	485
Db	406	DI---PVRRPESAAHVOELQEWLEGINSPKAREVLHTY	442
Sequence	456 AA;		
Query	Match	Score	
Best	Local	Similarity	
Matches	112;	Conservative	
Db	114	Mismatches	
QY	52	Indels	
Db	54	Gaps	
QY	107	Length	
Db	114		
QY	167		
Db	169		
QY	226		
Db	229		
QY	283		
Db	273		
QY	329		
Db	317		
QY	389		
Db	346		
QY	445		
Db	406		
Sequence	456 AA;		
Query	PGATTPKOLAGLRRGFDEFTLFGADLTMEGSELLRLTLEAHPRHDEPLPMPF	14.0%; Score 361.5; DB 22; Length 456;	
Matches	112; Conservative	24.3%; Pred. No. 27; Mismatches 174; Indels 99; Gaps 13;	
Db	114	V-TDASRRLCGFLKSLGVHVVFTTAADESFILESDRFTTAAESRERVRYRHSE--EERTLPLM	168
QY	52	VACAAAPAPATAPLSHVQQA-----LAELAKPKDDPRTKRVHQVAVAPVRAIAETGLA	106
Db	54	LACDSCTAAEGVQVLSQONAKDFRVLNLNKCDTSKHKVVLVSVCPQSLVYAAKFLS	113
QY	107	TRKQSEADRDNFCVADPTL--RQDHVITVVELGNIFKERGINLAELPEGEWDNPNGV	225
Db	169	TSACPGWVRYAERVLGRPITHLCTAKSPQOMGLAAMWKSYLAEKKGAKPDAMVNSIMPC	228
QY	226	YDKKLEALQE---SIPPALHOSRGACVLTSGEIAQIMEQDLSVRD-----	272
Db	283	GSGAGVLFG-----TTGGMEEALRTAVELFTGTPPLPRLSLEVRGMDCIK	328
QY	346	ANGLWNAKKLTQMAGEAKTDFVEIMACPACGCVGGGQPRS---TDKAITQRQALY	444
Db	406	AYGFRNIQNMILKKGKFPFHFVVELACAGCCCLNGRQAQTPDGHADKALIQRomegiya	405
QY	445	NLDEKSTLRRSHENPSPRELVTYLGPLGHKAELHHTY	485
Db	406	DI---PVRRPESAAHVOELQEWLEGINSPKAREVLHTY	442
Sequence	456 AA;		
Query	Match	Score	
Best	Local	Similarity	
Matches	112;	Conservative	
Db	114	Mismatches	
QY	52	Indels	
Db	54	Gaps	
QY	107	Length	
Db	114		
QY	167		
Db	169		
QY	226		
Db	229		
QY	283		
Db	273		
QY	329		
Db	317		
QY	389		
Db	346		
QY	445		
Db	406		
Sequence	456 AA;		
Query	Match	Score	
Best	Local	Similarity	
Matches	112;	Conservative	
Db	114	Mismatches	
QY	52	Indels	
Db	54	Gaps	
QY	107	Length	
Db	114		
QY	167		
Db	169		
QY	226		
Db	229		
QY	283		
Db	273		
QY	329		
Db	317		
QY	389		
Db	346		
QY	445		
Db	406		
Sequence	456 AA;		
Query	Match	Score	
Best	Local	Similarity	
Matches	112;	Conservative	
Db	114	Mismatches	
QY	52	Indels	
Db	54	Gaps	
QY	107	Length	
Db	114		
QY	167		
Db	169		
QY	226		
Db	229		
QY	283		
Db	273		
QY	329		
Db	317		
QY	389		
Db	346		
QY	445		
Db	406		
Sequence	456 AA;		
Query	Match	Score	
Best	Local	Similarity	
Matches	112;	Conservative	
Db	114	Mismatches	
QY	52	Indels	
Db	54	Gaps	
QY	107	Length	
Db	114		
QY	167		
Db	169		
QY	226		
Db	229		
QY	283		
Db	273		
QY	329		
Db	317		
QY	389		
Db	346		
QY	445		
Db	406		
Sequence	456 AA;		
Query	Match	Score	
Best	Local	Similarity	
Matches	112;	Conservative	
Db	114	Mismatches	
QY	52	Indels	
Db	54	Gaps	
QY	107	Length	
Db	114		
QY	167		
Db	169		
QY	226		
Db	229		
QY	283		
Db	273		
QY	329		
Db	317		
QY	389		
Db	346		
QY	445		
Db	406		
Sequence	456 AA;		
Query	Match	Score	
Best	Local	Similarity	
Matches	112;	Conservative	
Db	114	Mismatches	
QY	52	Indels	
Db	54	Gaps	
QY	107	Length	
Db	114		
QY	167		
Db	169		
QY	226		
Db	229		
QY	283		
Db	273		
QY	329		
Db	317		
QY	389		
Db	346		
QY	445		
Db	406		
Sequence	456 AA;		

CC as primers for polymerase chain reaction (PCR) amplification. GE
 CC proteins, their immunogenic fragments, and GE nucleic acid molecules
 CC encoding them are used to generate an immune response against GE,
 CC specifically as (genetic) vaccines, especially to control ehrlichiosis
 CC in humans and dogs, but also to raise Ab and to study DNA-protein
 CC interactions. Ab are used to detect GE proteins by forming an immune
 CC complex in standard assays, and correspondingly GE proteins can detect
 CC specific antibodies, especially for diagnosis, assessment and prognosis
 CC of GE infection, or of contamination of biological samples with GE.

QY 326 G---IKENITMPAPGSKFEELKKHRAAREAAAGTPGLAWDGAGFSEDGIGG 381

DB 4685 ERSRLIVENNDTARPVVISSSVPAFPQAAPDPAVAVGEG- VSWT---YRELDARSD 4739

QY 382 ITLVAVANGIGNAKKLTKMQ---AG---EAKDFVTEMA 416

DB 4740 ALARSLVLAAGVGVESPVVALERSPEVISAFLAVAKAG-CVFGVPUDSLHQARYDAV-AD 4798

QY 417 CPAG-----AALYNLDEKSTL 452

DB 4799 CGARTLAVALDRPMSQLTVVNSAGLGDSAVSGDUTDRAVVLPAAGPVPAVYRMTSST 4858

QY 453 RR---SHEN-----PSIRELYD----- 466

DB 4859 GRPKGVVTHQNLVLDATCTCGPPTPRVLFHAPHAFDASSYEWVPLLNGTVVAPRRS 4918

QY 467 ---TYLGEPGLGKAKHLLTHYWA 488

DB 4919 IDATVLRDIG-AHELTIVHVTAG 4941

RESULT 6

ID AAW89272

ID AAW89272 standard; Protein; 748 AA.

XX AAW89272;

AC AAW89272;

XX DT 04-MAR-1999 (first entry)

DE Granulocytic Ehrlichia protein clone s2.

XX KW Granulocytic ehrlichiosis; Ehrlichia sp.; GE protein; infection; tick; diagnosis; vaccine; antigenic protein; antibody; immune response.

XX OS Ehrlichia sp.

XX PN WO849313-A2.

XX PR 05-NOV-1998.

XX PF 24-APR-1998; 98WO-US08265.

XX PR 25-APR-1997; 97US-0044933.

XX (AQUI-) AQUILA BIOPHARMACEUTICALS INC.

XX PI Beltz GA, Coughlin RT, Murphy CA, Storey J;

XX WPI; 1999-009432/01..

DR N-PSDB; AAVB3787.

XX New nucleic acid from the human granulocytic ehrlichiosis agent - and related antigenic proteins; vectors, transformed cells and antibodies, useful for diagnosis and in protective vaccines

XX X Claim 16; Fig 7; 154pp; English.

XX The present sequence represents a granulocytic ehrlichia (GE) protein.

CC GE nucleic acids, vectors and host cells are used for the recombinant

CC production of GE proteins, and also in research to further characterise

CC the proteins. GE protein-encoding nucleic acid molecules are detected

CC by hybridisation to GE nucleic acid fragments or by using the fragments

Sequence 748 AA;

Query Match 4. 9%; Score 125.5; DB 20; Length 748;

Best Local Similarity 23.2%; Pred. No. 0.0055; Matches 66; Conservative 33; Mismatches 104; Indels 81; Gaps 12;

Db 98 LEHLITT-EVLSVNE---ITVPEKG---KTIILREALTSKGKVVKALIKNSADV 146

QY 308 FTGGLPLRSL-----SERGMDGKETHTITMVAPGS----- 340

Db 147 -NASPEPAITLGICGRCFCQGSKAIKHLKRVVEAGAHINTPTGSMSPLAAVQANAEASNL 205

QY 341 ---KFEELKKHRAAREAAAGTP-----GLPMLDGGFTSEGRGGT 363

Db 206 KEANNTVNFLHGRDLSSHTETPAPLHATAAGNHRITAMLIDKGAPATORDARGRTA 265

QY 384 LRVAVANGIGNAKKLTKMQAGEAKYDFVEIMAGPAGCAGVGGGGOPRSTDKAITOKRQAAL 443

Db 266 LHIANGDOOKLYMIAK-----KPDSC---OPLCSDMGDTALHE- AL 349

QY 444 Y-NLDEK---STURSHENPSIRLYDYLGEPLGKAKHLLH 482

Db 306 YSDINVTECFKLMMIKESRKLNSFFGDLINTPQEAENGDTLLH 349

RESULT 7

ID AAY78923

ID AAY78923 standard; Protein; 748 AA.

XX AC AAY78923;

XX DT 23-MAY-2000 (first entry)

DE Granulocytic ehrlichia NY1 protein sequence.

XX KW Granulocytic ehrlichia; granulocytic ehrlichiosis; vaccine; NY1; prevent; treatment.

XX OS Ehrlichia sp.

XX PN WO200006744-A1.

XX PD 10-FEB-2000.

XX PF 23-OCT-1998; 98WO-US22512.

XX PR 28-JUL-1998; 98US-0094381.

XX PA (AQUI-) AQUILA BIOPHARMACEUTICALS INC.

XX PI Murphy CI, Massung RF;

XX DR WPI; 2000-195304/17.

XX N-PRDB; AAZ92240.

XX Novel granulocytic ehrlichia nucleic acid molecules, their polypeptides

PT useful as vaccines for treating ehrlichiosis in mammals e.g. humans,

PT pigs and dogs -

XX Claim 15; Fig 24; 192pp; English.

CC This sequence represents the granulocytic ehrlichia (GE) NY1 protein
 CC sequence. The invention relates to 13 GE genes WI1, WI2, WI3, WI4, WIC,
 CC NY1, NY2, NY3, SWED, BOV, EQ, SLOV1, and SLOV2 isolated from 13 different
 CC GE clones from a dog, a cow, a horse and ten humans. Granulocytic
 CC ehrlichia is the causative agent of granulocytic ehrlichiosis, an acute
 CC potentially fatal tick-borne infection. A vaccine comprising a GE nucleic
 CC acid molecule or the polypeptide that it encodes, is used for producing
 CC an immune response in a host to prevent granulocytic ehrlichiosis in an
 CC animal. The protein sequences can be used to detect anti-GE antibodies in
 CC an animal.

XX Sequence 748 AA;

SQ Query Match 4.9%; Score 125.5; DB 21; Length 748;

Best Local Similarity 23.2%; Pred. No. 0.0055; Matches 66; Conservative 33; Mismatches 104; Indels 81; Gaps 12;

QY 248 LDHVITTEVGNIFVERGINLAELPEGEWDNPMSGVSGAVLFGTGGVNEAALRTAYEL 307
 98 LEHLITI-EVISVNEE----ITPPEGK----KTLTAEALTSGKVGVKALIRNSADY 146

Db 308 FTGTPUPRLSI-----SEVRGMDGJIKETNTIMVAPGS----- 340

QY 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349

RESULT 8'

AYY8924

XX standard; Protein; 748 AA.

AC AAY78924;

XX DT 23-MAY-2000 (first entry)

DE Granulocytic ehrlichia NY2 protein sequence.

XX KW Granulocytic ehrlichia; granulocytic ehrlichiosis; vaccine; NY2;

XX OS Ehrlichia sp.

XX PN WO200005744-A1.

XX PD 10-FEB-2000.

XX PF 23-OCT-1998; 98WO-US22512.

XX PR 28-JUL-1998; 98US-0094381.

XX PA (AQUI-) AQUILA BIOPHARMACEUTICALS INC.

XX PI Murphy CI, Massung RF;

XX DR N-PSDB; AAZ92241.

XX Novel granulocytic ehrlichia nucleic acid molecules, their polypeptides

PT useful as vaccines for treating ehrlichiosis in mammals e.g. humans,

PT pigs and dogs

XX Claim 15; Fig 25; 192PP; English.

PS

XX This sequence represents the granulocytic ehrlichia (GE) NY2 protein
 CC sequence. The invention relates to 13 GE genes WI1, WI2, WI3, WI4, WIC,
 CC NY1, NY2, NY3, SWED, BOV, EQ, SLOV1, and SLOV2 isolated from 13 different
 CC GE clones from a dog, a cow, a horse and ten humans. Granulocytic
 CC ehrlichia is the causative agent of granulocytic ehrlichiosis, an acute
 CC potentially fatal tick-borne infection. A vaccine comprising a GE nucleic
 CC acid molecule or the polypeptide that it encodes, is used for producing
 CC an immune response in a host to prevent granulocytic ehrlichiosis in an
 CC animal. The protein sequences can be used to detect anti-GE antibodies in
 CC an animal.

XX Sequence 748 AA;

SQ Query Match 4.9%; Score 125.5; DB 21; Length 748;

Best Local Similarity 23.2%; Pred. No. 0.0055; Matches 66; Conservative 33; Mismatches 104; Indels 81; Gaps 12;

QY 248 LDHVITTEVGNIFVERGINLAELPEGEWDNPMSGVSGAVLFGTGGVNEAALRTAYEL 307
 98 LEHLITI-EVISVNEE----ITPPEGK----KTLTAEALTSGKVGVKALIRNSADY 146

Db 308 FTGTPUPRLSI-----SEVRGMDGJIKETNTIMVAPGS----- 340

QY 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349Db 147 -NASPBATIGIQCFCGSKAITHLKVVEAGHINTPTGSHSPLAAVQANEASNL 205
 341 ---KFEELIKHRAARAEEAAGHGP-----GPLANDGGAGFTSEDGRGGT 383

Db 206 KEANKIVNLFHLRGADLSSEHTGPALHATAAGNHRPMLLDKGAPATQRDARGRA 265

QY 384 LRVAVANGLGNNAKKLITKMOAGEAKYDFVEMACPGCVGGGSPRSTKAIKQRQAL 443
 266 LHTAANGDCKLYRMIAK-----KOPDSC----QPLCSIDMGDTALHE-AL 305Db 444 Y--NIDEK--STLRSHENPSPIRELYDTYLGEPGHKAELLH 482
 306 YSDNTEKCFIKMLKESRKHLNSSFFGDLNTPOEANGDTLLH 349

PI Spaderna SK, Quinn KE, Shimkets RA, Muralidhara P, Spytek KA;
 XX DR WPI: 2001-398154/42.
 DR N-FSDB; AAS00333.
 PT Novel polypeptide comprising members of protein families (e.g., seven-pass transmembrane receptor proteins) according to presence of domains and sequence relatedness are useful for treating or preventing, e.g., Alzheimer's and Parkinson's -

PS XX Claim 1; Fig 8; 162pp; English.

CC The sequence represents the amino acid sequence of glutamate receptor-like protein, MEM2, selected from a group (MEM1-MEM8) comprising members of protein families according to the presence of domains and sequence relatedness, e.g., seven-pass transmembrane receptor protein (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein (MEM5), phosphatase I protein (MEM6), and retinol-binding protein (MEM7-MEM8). The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are all useful for treating or preventing a pathology associated with (I) comprising administering (I), (II), or (III) to a subject (preferably a human). In addition, (I), (II), and (III) may be used to manufacture a medicament for treating a syndrome associated with a human disease that is associated with (I). Furthermore, (I) may be used to identify agents that bind to it, screen modulators of its activity and determine the presence or predisposition to disease associated with altered levels of (I). Disorders for MEM1 include Alzheimer's or Parkinson's Disease, cancer, nephrology, and female reproductive health. Disorders for MEM4 include those involving the lung and/or brain (e.g., schizophrenia). For MEM5, disorders include heart (arrhythmic disorders) and other muscular disorders, clotting deficiencies and cobalamin deficiencies (e.g., pernicious anaemia). Such disorders for MEM6 include diabetes, whereas disorders for MEM7 and MEM8 include vision-related disorders, cancer, and other neoplastic pathologies.

SQ Sequence 965 AA:

Query Match 4.7%; Score 122; DB 22; Length 965; Best Local Similarity 22.3%; Pred. No. 0.019; Matches 172; Conservative 45; Mismatches 187; Indels 194; Gaps 23;

QY 7 KPCAAVSIKGSSCRAQVAPRPLAATSVRVALTLEAPARRIG--NVAACAAAAPAAEA 63

25 QPGCVLARLGGSVRLGALIPRPLARALARALAR-ANALPRPHXLSLELVVAAAPPAD 83

QY 64 PLSHVO--QAL----AELAKPKD--TRKIVCVQVAPAVRAETLGAP-GATI 111

84 PASLTRLGLQALVPGVAAVLLAFAEPPRLQHFLAATETPVLSLREARAPXGAPN 143

112 PKQLAEGLRIGFEDVEFDLFGADLTIMEGSELLHRLEHNPNSDPLMFTSCCP 171

144 PFHL----OLHWASPLT-----LIDIVAVLORHAWEDVGLACRQDP 184

QY 172 GWIAMEKSY---PDLPLIVVSSCKSPQMLAAMVKSIAEKGKGIAPKDMVMSIMPCTR 227

185 GGLVALWTSRAGRGRPPOLVLDLSXRDTGQGLRNLAPMAPVGEAP----- 231

Db 228 K0SEADRDWFCVDADPTLQLDHVITTVLG-NIFKERCINABLEPEGEWDNPMSGVSGA 286

Db 232 -----VPAAVLILGCDIARARRVLEAVPPGHW----- 258

QY 287 GVLFGTTGCVMEALRTAYELEFTGTPPLRPLSLSEVR-----GMGDT 327

Db 259 --LKGTP--LPPKALPTA----GLPXXGLLALGSEVARPPLAATHDIVQVVARALGSA 309

QY 328 KEINTIMVPA-----GSKFSEBLKIRAAARAAEAAAGTGPGLAWDGAGTSEDRG 380

310 VQPKRALLKAPVNGDLOPAGPSPGFLARFLANTXFGRTGPV-WTGSS-PDEGQ- 366

QY 381 GITLRLVAVANGLGNAKKLTKMGAGEAKYDFVEIMACAG--CVGGGQQPRSTKATQK 438

Db 367 -----CPAGQCL---DPGTNDSATLDA 386

QY RESULT 12
 AAU02198 ID AAU02198 standard; Protein: 971 AA.
 XX ID AAU02198:
 XX AC 26-SEP-2001 (first entry)
 XX DT Human glutamate receptor-like protein, MEM3.
 XX DE Glutamate receptor; MEM3; therapeutic; diagnostic; MEM2;
 KW human; Alzheimer's disease; Parkinson's disease; cancer; nephrology;
 KW female reproductive health; lung disorder; brain disorder; schizophrenia;
 KW heart disorder; arrhythmia; muscular disorder; clotting deficiency; MEM3;
 KW cobalamin deficiency; pernicious anaemia; diabetes; MEM4; MEM5; MEM6;
 KW vision-related disorder; neoplastic pathology; MEM7; MEM8.
 XX OS Homo sapiens.
 XX PN WO20014473-A2.
 XX PD 21-JUN-2001.
 XX PR 14-DEC-2000; 2000MO-US33909.
 XX PR 14-DEC-1999; 9905-0170564.
 PR 27-DEC-1999; 9905-0173362.
 PR 29-DEC-1999; 9905-0173544.
 PR 04-JAN-2000; 2000US-9966564.
 PR 09-AUG-2000; 2000US-0223929.
 PR 13-DEC-2000; 2000US-9966565.
 XX PA (CIBA-) CURAGEN CORP.
 XX PI Spaderna SK, Quinn KE, Shimkets RA, Muralidhara P, Spytek KA;
 XX DR WPI; 2001-398154/42.
 XX N-FSDB; AAS00334.

PS XX Claim 1; Fig 13; 162pp; English.

CC The sequence represents the amino acid sequence of glutamate receptor-like protein, MEM3, selected from a group (MEM1-MEM8) comprising members of protein families according to the presence of domains and sequence relatedness, e.g., seven-pass transmembrane receptor protein (MEM1), glutamate receptor (MEM2-MEM4), potassium channel protein (MEM5), phosphatase I protein (MEM6), and retinol-binding protein (MEM7-MEM8). The MEM polypeptides (I), nucleic acids (II), and antibodies (III) are all useful for treating or preventing a pathology associated with (I) comprising administering (I), (II), or (III) to a subject (preferably a human). In addition, (I), (II), and (III) may be used to manufacture a medicament for treating a syndrome associated with a human disease that is associated with (I). Furthermore, (I) may be used to identify agents that bind to it, screen modulators of its activity and determine the presence or predisposition to a disease associated with altered levels of (I). Disorders for MEM1 include Alzheimer's or Parkinson's Disease, cancer, nephrology, and female reproductive health. Disorders for MEM4 include those involving the lung and/or brain (e.g., schizophrenia). For

XX
DR
N-PSDB; AAI4781.

PT Novel isolated ramoplanin biosynthetic pathway polypeptide useful for
PT chemically modifying biological molecule that is a substrate for a
PT polypeptide encoded by a ramoplanin biosynthesis gene cluster
XX

Claim 14; Page 169-186; 212pp; English.
XX
CC The invention relates to an isolated ramoplanin biosynthetic pathway
CC polypeptide selected from a polypeptide of open reading frames (ORF)
CC 1-32. The isolated polypeptides are useful for chemically modifying a
CC biological molecule that is a substrate for a polypeptide encoded by a
CC ramoplanin biosynthesis gene cluster, by contacting the biological
CC molecule with the isolated polypeptide, where the polypeptide chemically
CC modifies the biological molecule. The method comprises contacting the
CC biological molecule with at least two different polypeptides encoded by
CC ramoplanin ORFs 1-31. The polypeptides are useful for directing the
CC biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated
CC gene cluster comprising the ORFs is useful as a substrate for
CC bioengineering of antibiotic structures. An isolated polypeptide or its
CC encoding nucleic acid sequence is useful for generating derivatives of
CC ramoplanin, for improving production or for producing variants of other
CC antibiotics of the peptide class. The isolated polypeptides are useful
CC for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain
CC in conjunction with other peptide synthetase modules and allowing the
CC incorporation of Threonine into a peptide antibiotic precursor, for modifying
CC fatty acid structure and/or enhancing fatty acid incorporation into the
CC peptide antibiotic structure, for production of an hydroxyphenylglycine
CC (HPG)-containing peptide antibiotic, for enhancing secretion of
CC ramoplanin or its variants and derivatives, for enhancing uptake of
CC precursors for ramoplanin biosynthesis, for enhancing production of
CC ramoplanin products or its variants or derivatives, to chlorinate HPG of
CC a peptide antibiotic precursor, and for designing specific nucleotide
CC probes and primers for identifying and isolating putative lidepeptides
CC -producing microorganisms. This sequence represents one of the ORF
CC proteins of the ramoplanin producing *Actinoplanes* sp. microorganism of
CC the invention.

Sequence 4999 AA;

Query Match 4.5%; Score 116; DB 23; Length 4999;

Best Local Similarity 24.1%; Pred. No. 1.1; Matches 100; Conservative 32; Mismatches 171; Indels 112; Gaps 20;

QY 8 PCAAVSISGSSCRAQVA--PRAPLAMSTVRVALATEAPARRLGNWACAMAPAAEAP 64
DP 44400 PDSALISQVAHWR-RQLLAGAPDELPLPDPHPRAEATR-----GHVVFVPPA--- 4448
QY 65 LSHVQOQALAEALAKPKDDPTRKHYCVQVAVRAVIAETIG-LAPGATTPKQLAEGLRRLUG 123
DP 44419 --VHHQIAELARRNG-----VTVFMTVQATALAVILSKLGAGTDIPIGVAVAGRT-- 4495
QY 124 FDEVFDTIPLG-----ADUTIMEBESSELLHLRLTE-HLEAHPHSDDEPLPMTSCCGW 173
DP 44496 -DPTLDNLIGFFVNTLVRDITDLDLHRTDPTLHAFTHDQPV----- 4543
QY 174 IAMELKSPDILIPVSSKSP-OMMLAAMVSKYLAEKKGIAFKRDMVAVSIMPCTRKSE 231
DP 4544 --FEKULDAPLPIRSRHLRPLQMMQLQSIGRAGEAELPQIETAVLSPGGVAAK-- 4597
QY 232 ADRWIFCYDADPTLRLQDHVITVFLGNIKFERGINLAELPEGEWDNIMGVGSGAGVIFG 291
DP 4598 -----VLDL-----LSLSAYDDG-----RPAGL--AGTIVVA 4623
QY 292 TTG---GMEALRTRATELFITGPL-PRLSLSBVRGKG-----IKETNTITWPABGS 340
DP 4624 AADLFEGHFGPAAERTAGYTLARLAVLPADGARLGVDLGEERLVLGWNHDTAAVPAV 4683
QY 341 KFEFLKLKRAARAEEAAMGATPGPLAWDGGA---GFTSDEGRGGITLRAVANGLG 393
DP 4684 AVPELIERRAAAPERGA-----VWCGDTHLRYGEELNARANRNLARLVERGAG 4731

Search completed: June 3, 2003, 16:07:35
Job time : 75 secs

This Page Blank (uspto)

A;Residues: 1-606 <STO>
 A;Cross-references: EMBL:X57838; NID:940827; PIDN:CAA40970.1; PID:940828
 C;Genetics:
 A;Gene: hydc
 C;Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology
 C;Keywords: 2Fe-2S; 4Fe-4S; iron; metalloprotein; oxidoreductase
 F;33,45,48,66/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted
 F;98,102,105,111/Binding site: 4Fe-4S cluster (His, Cys, Cys, Cys) (covalent) #status predicted
 F;152,155,158,205/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
 F;162,195,198,201/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
 F;305,360,507,511/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
 F;511/Binding site: diiron cofactor (Cys) #status predicted

Query Match 30.8%; Score 794.5; DB 2; Length 606;

Best Local Similarity 42.4%; Pred. No. 3.1e-48; Indels 57; Gaps 6;
 Matches 169; Conservative 61; Mismatches 112; Indels 57; Gaps 6;

QY 90 QVAPAVRVAIAETGLGLARCATTPKOLAECLRLRGDFEVDFDTLGRADLTIMEEGSELLIRL 149

Db 233 QFAPAVRVLGEERGLGPLPSVSVQVPAALRLLGADVWDTNFRADLVMEEGTELLQL 292

QY 150 TEHIEAHPHSDEPLPMTFSCCPGMIAMLEKSPOLIPYVSSCKSPQMLAAMVSKYALEK 209

Db 293 -----RGAKAKPLFTSCCPGMIAMLEKSPOLIPYVSSCKSPQMLAAMVSKYALEK 344

QY 210 KGTIAPKDMVAVSIMPCTKQSEADRWDICVADPDLQLDHVITVLEGNIKFERGINA 269

Db 345 MNVAPERMRVYSLMPCTKARKEEARPERFRDG--VRQDAVLTREFAURLRREGIOLA 401

QY 270 ELPGEWHDNP-MGVSAGSAGVLFGTTGGVMEALRTAYELFTGTLPRLPRLSLEVRGMGDGK 328

Db 402 GLEPSPCDPDLIMGRMAGVAVIFGTTGGVMEALRKVYHVYLNGKELAPVHLALKYENR 461

QY 329 ETNTIMVPPAGSKFEELLKRAAARAEEAHGTGFLAMGGAGFTSEDEGGITLRAV 388

Db 462 EAVV-----PL-----GEGNSVYKAV 478

QY 389 ANGLGNNAKKLITKMQAGAEKAYDFVIMACPAGCVGGGQPRS--TDKAYTOKRQALYNL 446

Db 479 VHGLKAARQMEAVLAGKADHVFVEMACPGGGQGPRSKRQAYNPAQARRAFLS 538

QY 447 DEKSPLRSHENSPIRELDTYLGEPHKAHELIHTHY 485

Db 539 DAENALROSHNNPLIGKIVTESFLGEPCCNSLNSHLRHLHTY 577

RESULT 5
 HQC1P

hydrogenase (EC 1.18.99.1) (Fe) I, periplasmic [validated] - Clostridium pasteurianum
 N;Alternate names: [Fe] hydrogenase
 C;Species: Clostridium pasteurianum

C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 15-Sep-2000
 C;Accession: A40330

R;Meyer, J.; Gagnon, J.

Biotechnology 30, 9697-9704, 1991
 Article: Primary structure of hydrogenase I from Clostridium pasteurianum.

A;Reference number: A40330; MUID:92002005; PMID:1911757

A;Accession: A40330

A;Molecule type: DNA

A;Residues: 1-574 <MEY>

A;Cross-references: GB:M81737; GB:M62754; NID:9144835; PIDN:AA23248.1; PID:9144836

R;Peters, J.W.; Lanzilotta, W.N.; Lemon, B.J.; Seefeldt, L.C.

Science 282, 1853-1858, 1998

A;Title: X-ray crystal structure of the Fe-only hydrogenase (CpI) from Clostridium past

A;Reference number: A59203; MUID:9905388; PMID:9836629

A;Contents: annotation; X-ray crystallography, 1.80 angstroms, residues 1-574

A;Contents: annotation; X-ray crystallography, 1.80 angstroms

R;Peters, J.W.; Lanzilotta, W.N.; Lemon, B.J.; Seefeldt, L.C.

Science 282, 1853-1858, 1998

A;Reference number: A77988; PDB:4FEH

A;Comments: annotation; X-ray crystallography, 1.80 angstroms, residues 1-574

C;Comment: Three distinct hydrogenases, the Fe, NiFe, and NiFeSe hydrogenases, have been

I), each consisting of a single polypeptide chain.

C;Function: diiron cofactor (Cys) #status predicted

RESULT 6

JC6002

hydrogenase (EC 1.18.99.1) (Fe) large chain [similarity] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 19-May-2000

C;Accession: JC6002

R;Gorwa, M.F.; Croux, C.; Soucaille, P.

J. Bacteriol. 178, 2668-2675, 1996

A;Title: Molecular characterization and transcriptional analysis of the putative hyd

A;Reference number: A59202; MUID:96196176; PMID:8626337

A;Accession: JC6002

A;Molecule type: DNA

A;Residues: 1-582 <GOR>

A;Cross-references: GB:U15277; NID:957063; PIDN:AAB0373.1; PID:9557064

A;Experimental source: ATCC824

C;Genetics:

A;Gene: hydc

C;Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology

C;Keywords: 2Fe-2S; 4Fe-4S; iron; metalloprotein; oxidoreductase

F;139,207/Domain: ferredoxin 2[4Fe-4S] homology <FER>

F;34,45,48,61/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

F;93,97,100,106/Binding site: 4Fe-4S cluster (His, Cys, Cys, Cys) (covalent) #status predicted

F;146,149,152,195/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F;166,189,192,195/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F;299,354,498,502/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

A;Description: catalyzes the reduction of two protons to molecular hydrogen by reducing NADH. It is involved in the reduction of NADP to NAD.

C;Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology

C;Keywords: 2Fe-2S; 4Fe-4S; hydrogen metabolism; iron-sulfur protein; metalloprotein,

F;140-208/Domain: ferredoxin 2[4Fe-4S] homology <FER>

F;34,46,49,62/Binding site: 2Fe-2S cluster (Cys) (covalent) #status experimental

F;94,98,101,107/Binding site: 4Fe-4S cluster (His, Cys, Cys, Cys) (covalent) #status experimental

F;147,150,153,200/Binding site: 4Fe-4S cluster (Cys) (covalent) #status experimental

F;157,190,193,196/Binding site: 4Fe-4S cluster (Cys) (covalent) #status experimental

F;503/Binding site: diiron cofactor (Cys) #status experimental

Query Match 28.9%; Score 743.5; DB 1; Length 574;
 Best Local Similarity 37.1%; Pred. No. 1.1e-44; Indels 69; Gaps 6;
 Matches 163; Conservative 65; Mismatches 112; Indels 69; Gaps 6;

QY 52 VACAAAPAAEPLSHVQAAELAKPKDPDTRKHCYQVAPAVRVAETGLGAPGATT 111

Db 198 IACPVAAEPLSHVQAAELAKPKDPDTRKHCYQVAPAVRVAETGLGAPGATT 249

QY 112 PKQIAEGLRRGFDEDFDTLGRADLTIMEBGSELHLRTEHLAKPHSDEPLPMTSCCP 171

Db 250 TGKIXTALROLGFDFKIDINFAGAMTIEATEVQI-----ENGPFPFTSCCP 301

Db 172 GWIAMELEKSYPDLLIVYSSCKSPQMLAAMVSKYLAEGKIGIAPKDMVWSIMPCTKOE 231

Db 302 GWVROAENYYPELUNLSSAKSPQOIFGTTASKTYPSTSGLDPKVNFTVMPCTKFE 361

Db 232 ADPDMCVDADPDLRQLDHVITVLEGNIKFERGINAHLPEGMDNPAGVSGAVLG 291

Db 362 ADPOMEKG---DADIVATLTTTAKLMDAKIPEAKLSDRPAKEDSPLRNGEYSAGA 418

Db 292 TTGGMEALRTAYELFTGTLPRLPRLSLEVRGMODGKINETNIMDPAPGKFEELKHRAA 351

Db 419 ATGGMEALRSAKDFAENAELEDEIYKQVRGLNGIKEVEI--NNKKY----- 466

Db 352 ARAEAAGHGPGLPGLWDGAGFTSDGREGITLRAVANGLGNRKLTKMQACRKYD 411

Db 467 -----INVAVINGASNLKFKMGSQMINKQYHF 493

Db 412 VEIMACPAGCVGGGQPRSDFK-----ATQKRAALYNIDEKTLRSHENPSIRELYD 465

Db 494 TAVMGCGSCVNGGQSPHVKPDKLDRVDTKVRASVLYQDEHLSKRKHENTLVLKMTQ 553

Db 467 TYLGPDLGHKAELIHTHY 485

Db 554 NYFGKRGEGRAEHLHFY 572

F; 502/Binding site: dilon cofactor (Cys) #status predicted

Query Match 28.8%; Score 741; DB 2; Length 582;

Best Local Similarity 37.9%; Pred. No. 1.8e-44; Indels 72; Gaps 9;

Matches 167; Conservative 68; Mismatches 134; Delins 72; Gaps 9;

Qy 52 VACAAAPAEAPLISHQOALAEELAKPDKDPTRKHCVQVAPAVRVAETIGLAPGATP 111

Db 197 TACPVAALEKSHIEKVOEAL-----NDP-KKHVIVAMAPSRTAKMELFKMGYGVKU 248

Qy 112 PKQAEGLERLGFDEVFTLFGADLTMEGSEELHRTEHLEAHPHSDEPLPMTSCCP 171

Db 249 TGKLYTALLMLGEDKVFDSNFGADMTEATELGRV-----KNGPFFMFTSCCP 300

Qy 172 GWIAMELEKSPDLPYVSSCKSPQMLAAVSKYLAEKKGIAPKDMVMSIMPCNKKYE 360

Db 301 AWVLAQNYHPELLDNSSAKSPQOIFGASKTIPSISSGIAPEVDYVTPIMONDKKE 360

Qy 232 ADROWFCVADPTLRQDIDVITYELGNFFKERCINLALPPEGWDNPQGAGVLF 291

Db 361 ADIPFMEV-----SLRDIDASLTLRELAKMKDAKIKFADLEGEVDPAMGTYSGAGAIFG 417

Qy 361 ADIPFMEV-----SLRDIDASLTLRELAKMKDAKIKFADLEGEVDPAMGTYSGAGAIFG 417

Db 292 TTGGMEALRTAVELFTGPLRSLVRSYRGMGKINETITWVAPGSKFEELKHRAA 351

Qy 418 ATGGVMEALRSAKDFAENKELENVDTYEVGRGKIKEAEV----- 460

Db 432 VEIMACPACVGSSGQP--RSTDKA--ITOKRQAAVLNDESTL--RRSHENPSIHL 464

Db 493 VEIMACPACVGSSGQP--RSTDKA--ITOKRQAAVLNDESTL--RRSHENPSIHL 464

Qy 493 IEVMACPGCINGGGQPHNADRENVDYRKLRASVLYNQD-KVVLSKRKSHDNPAIKM 551

Qy 465 YDTVIGEPLCHKAELLHRY 485

Db 552 YDSYFGKPGEGLAHKLLHVY 572

Db 572 YDSYFGKPGEGLAHKLLHVY 572

RESULT 7

D9503 hydrogene dehydrogenase [imported] - clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001

C;Accession: D9503

R;Nelling, J.; Breton, G.; Omelchenko, M.V.; Marharova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-producing Bacterium C. C;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: D96903

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-582 <KUR>

A;Cross-References: GB:X00437; PIDN:AAK78015.1; PIDN:915022848; GSPDB:GN00168

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC0028

C;Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology

Query Match 28.8%; Score 741; DB 2; Length 582;

Best Local Similarity 37.9%; Pred. No. 1.8e-44; Indels 72; Gaps 9;

Matches 167; Conservative 68; Mismatches 134; Delins 72; Gaps 9;

Qy 52 VACAAAPAEAPLISHQOALAEELAKPDKDPTRKHCVQVAPAVRVAETIGLAPGATP 111

Db 197 TACPVAALEKSHIEKVOEAL-----NDP-KKHVIVAMAPSRTAKMELFKMGYGVKU 248

Qy 112 PKQAEGLERLGFDEVFTLFGADLTMEGSEELHRTEHLEAHPHSDEPLPMTSCCP 171

Db 249 TGKLYTALLMLGEDKVFDSNFGADMTEATELGRV-----KNGPFFMFTSCCP 300

Qy 172 GWIAMELEKSPDLPYVSSCKSPQMLAAVSKYLAEKKGIAPKDMVMSIMPCNKKYE 360

Db 301 AWVLAQNYHPELLDNSSAKSPQOIFGASKTIPSISSGIAPEVDYVTPIMONDKKE 360

Db 323 ADROWFCVADPTLRQDIDVITYELGNFFKERCINLALPPEGWDNPQGAGVLF 291

Db 361 ADIPFMEV-----SLRDIDASLTLRELAKMKDAKIKFADLEGEVDPAMGTYSGAGAIFG 417

Qy 292 TTGGMEALRTAVELFTGPLRSLVRSYRGMGKINETITWVAPGSKFEELKHRAA 351

Db 418 ATGGVMEALRSAKDFAENKELENVDTYEVGRGKIKEAEV----- 460

Db 461 -----AGNKLNAVINGASNEFFMSKGKMKNEQHF 492

Qy 412 VEIMACPACVGSSGQP--RSTDKA--ITOKRQAAVLNDESTL--RRSHENPSIHL 464

Db 493 VEIMACPACVGSSGQP--RSTDKA--ITOKRQAAVLNDESTL--RRSHENPSIHL 464

Qy 493 VEIMACPGCINGGGQPHNADRENVDYRKLRASVLYNQD-KVVLSKRKSHDNPAIKM 551

Db 552 YDSYFGKPGEGLAHKLLHVY 572

Db 572 YDSYFGKPGEGLAHKLLHVY 572

RESULT 8

HQDVFLL

cytochrome-c3 hydrogenase (EC 1.12.2.1) (Fe) large chain - Desulfovibrio vulgaris (st N;Alternate names: hydrogenase (Fe) alpha chain

C;Species: Desulfovibrio vulgaris

C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 05-May-2000

C;Accession: A24551; B27480

R;Voodenaw, G.; Brenner, S.

Eur. J. Biochem. 148, 515-520, 1985

A;Title: Nucleotide sequence of the gene encoding the hydrogenase from Desulfovibrio A;Reference number: A24551; MUID:85203856; PMID:388621

A;Accession: A24551

A;Molecule type: DNA

A;Residues: 1-421 <VOO>

A;Cross-References: GB:X02416; NID:940829; PIDN:CAA26266.1; PID:940830

B;Pritchill, B.C.; He, S.H.; Li, C.; Menon, N.; Choi, E.S.; Przybyla, A.E.; Darvartani Biochem. Biophys. Res. Commun. 149, 369-377, 1987

A;Title: Identification of three classes of hydrogenase in the genus, Desulfovibrio. A;Reference number: A27480; MUID:88106446; PMID:332275

A;Accession: B27480

A;Molecule type: protein

A;Residues: 1-34; K <PR>

C;Comment: Hydrogenases catalyze reactions involving the production or consumption of H₂, N₂Fe, and N₂FeS₂ hydrogenases, are found in the periplasm of sulfate-reducing bacteria. The active D. vulgaris hydrogenase (Fe) is a dimer of large 'alpha' and small 'beta' subunits. The active D. vulgaris hydrogenase, are found in the periplasm of sulfate-reducing bacteria. The active D. vulgaris hydrogenase (Fe) is a dimer of large 'alpha' and small 'beta' subunits. C;Genetics:

A;Gene: hydA

C;Superfamily: hydrogenase (Fe) large chain; ferredoxin 2[4Fe-4S] homology

C;Keywords: 4Fe-4S; hydrogen metabolism; iron-sulfur protein; metalloprotein; oxidore F;28.84/Domain: ferredoxin 2[4Fe-4S] homology <PR> F;35.38.41,76/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted F;45.66,69,72/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted F;382/Binding site: dilon cofactor (Cys) #status predicted

Query Match 24.6%; Score 633; DB 1; Length 421;

Best Local Similarity 36.6%; Pred. No. 4.7e-37; Matches 149; Conservative 56; Mismatches 122; Indels 80; Gaps 9;

Qy 54 CAAAPAEAPLISHQOALAEELAKPDKDPTRKHCVQVAPAVRVAETIGLAPGATP 112

Db 72 CLTHOP--ENAYEWSWPEVERKLKDGVK-CIAMPAPAVRYALGAFGMGVSVT 127

Qy 113 KOLAGLRLGFDEVFTLFGADLTMEGSEELHRTEHLEAHPHSDEPLPMTSCCP 172

Db 128 GKMALQQLGFAHWDTEFTADVIWEESEFVERLTK-----KSDMPLPMTSCCP 181

Qy 173 WIAMELEKSPDLPYVSSCKSPQMLAAVSKYLAEKKGIAPKDMVMSIMPCNKKYE 323

hypothetical protein SPCC1450.10c - fission yeast (Schizosaccharomyces pombe)
 C;Species: Schizosaccharomyces pombe
 C;Accession: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C;Reference number: Z21962
 A;Accession: T40922
 A;Status: preliminary; translated from GB/EMBL/DDJB
 A;Molecule type: DNA
 A;Gene: SPDB:SPCC1450.10c
 A;Map position: 3
 A;INTRONS: 31/3

Query Match 12.1%; Score 311.5; DB 2; Length 538;
 Best Local Similarity 24.2%; Pred. No. 3e-14;
 Matches 123; Conservative 84; Mismatches 179; Indels 123; Gaps 19;

Qy 52 VACAAAPAAEAPSHQVQOALAKPKDPDKTRKIVCVOVAPAVVIAETLGLARGATT 111
 65 LACSGCITSATEVVLNQSQYQEVLRKHLERSKQELIVYSLSPQVANLAAVYGLSQEIQ 124

Qy 112 PKQLEGRLAGEDFVDTLEGADLTIMEGSEELLRLTHLEAH----- 156

Qy 125 AVLERAVPIGKLGFLAIDTNASREIVLQOCQAEFCNSWLS - RAHKNQONTNSVNEHP 183

Qy 157 --PHS-----DEPLPMETSCCPGWIAMELEKSPDLPVPSCKSPQ 195

Qy 184 LIPHSTSQISGVHSNTSSNGNINERAVLPILSSPGCIVYEVKTHSNLIPNLSVRSPQ 243

Qy 196 RMLAMAVKSLAEGKJAPKQMVMSIMPCTRKQBADRWFCVADPTLRQLDHWITV 255

Db 244 QACGRILKDWAVQFSMQRNDWHLISLMPCEFDKKELEASREF--SENGVRDVDSLTPK 300

Qy 256 ELMNTEKFERGGINLAEPEGEWENDPMP-----GVGSAGAVLGTGTTGG 295

Qy 301 ELVEMRFLRIDPILTRK---NPPIFQOSTDAIKFPPWPRITYEQIGS-----SSGG 349

Db 296 VMEALRATVLEFTGP--LPLRLSSEVRKGDKIG-----ETN--ITMVPAPSKFEE 344

Qy 350 YMGYYVLSYAAKMLRGIDDVGPYVSHNKKNG-DLTETYLHPETNPQLISMATCYGFRNQ 408

Qy 345 LKKHRARAAEAAATGPGPLAWDAGFTSDEGDRGIGTAVANGLGNARKLITRQMA 404

Db 409 NLVRR-----VHG-----NSSVRKGRVLLKKRVRNSQNPTE----- 440

Qy 405 GEAKYDFVEIMACPAGCAGVGGGO-PRSTDKRAITOKRQ---AALYNDKSTLRSHEN 458

Qy 441 EPRSDIVEVFACPGCGCINGGQQLPPSVRIVSARDWQJOVERKLY--EPPG--RSVDQ 496

Qy 459 PSIREKDYTVIGEPGKHKAHELLHHYVA 487

Qy 497 SAVSYMLEQWVKDPT--LTPKFLHTSYRA 523

RESULT 12

S63306 LET1 protein homolog YNL240C - yeast (Saccharomyces cerevisiae)
 N;Alternative names: hypothetical protein N114
 C;Species: Saccharomyces cerevisiae
 C;Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
 C;Accession: S63206; JQ0891; S67338; S72076
 R;Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.
 A;Submitted to the Protein Sequence Database, April 1996
 A;Reference number: S63188
 A;Accession: S63206

Query Match 9.9%; Score 255; DB 2; Length 491;
 Best Local Similarity 23.4%; Pred. No. 2.6e-10;
 Matches 114; Conservative 72; Mismatches 189; Indels 118; Gaps 18;

Qy 52 VACAAAPAAEAPSHQVQ------QALAEALKPKDPDKTRKIVCVOVAPAVVIAETLGL 105

Db 60 LACSGCITSSEETLSSQSHSVYELKNGKLSQOD-----KFLWVTSQPCRLSAGYGL 115

Qy 106 APGATTPKQLABGJRLRGDFEVDTLRFQADLTIME-----EGSEL-----LHRL 149

Db 116 -----TLEADCLMNFFQKHFQCKYMGVTEKGRISIKT 151

Qy 150 TEHLEAPHSDPDL-----PMFTSCCPGWIAMELEKSPDLPVPSCKSPDNMLAAMKS 204

Db 152 VETIAHKHQKENTGADRKPPLSACVCPGFLTYTEKTPQVLVPMILNVKSPQITGSLRA 211

Qy 205 YLAEGKJAPKQMVMSIMPCTRKQBADRWFCVADPTLRQLDHWITVLEGNIFER 264

Db 212 TPE-ESLAIARESEFYHSLMPCEFDKPLASPRSPSDDG-----IDCVITPRELVTMQEL 264

Qy 265 GINLAEI-----PECEWDNPMPGVMGGAGVLFGTGCGVMEALRATVLEFTGTT- 311

Db 265 NIDFKSFLTEDTSLYGLRSPG-WDPVWHWASNLG--GTCGGY-----AQYQTVAVQ 313

Qy 312 ---PLPLSLSEVGRGMGIKENITMPAPPSKFEELIKR--AAARAEMAAHGTGPPL 365

Db 314 RLHPSOMIVLEGGRNSDITV-----EYRLHDIDRIAAASELSGFRNQNL 358

Qy 366 AMDGAGFTSDEGRRGTTLRAVANGLGNARKLITKMQAGEAKY--DFVEIMACPAGCAGV 423

Db 359 VRKLTSGSGSERKRNTRKRTGPKNSREMAAATADPYHSIDYIEVNACPGACMN 418

Qy 424 GCGQPRSTDKATOKRQALYNLNDEKSTLRSHENPSIRELVDTYGEPLGKHKAHELLHT 483

Db 419 GCGLNLGEQNSLKRKQLV-----OTLNKKH-GEELAMVQPLTGPKLEAAARPLSL 469

Qy 484 HYVAGGYEEKBDK 496

Db 470 EVFAPVQQAVER 482

RESULT 13

B97297

hydrogenase chain (ferredoxin) [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C;Accession: B97297

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Kocon, E.V.; Smith, D.R.

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A;Cross-references: GB:AE001437; PIDN:AAK81165.1; PID:915026303; GSPDB:GN00168

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Status: Preliminary

A;Molecule type: DNA

A;Residues: 1-450 <KUR>

A;Experimental source: Clostridium acetobutylicum AACC824

A;Gene: CAC3230

Query Match 9.5%; Score 245.5; DB 2; Length 450;

Best Local Similarity 24.6%; Pred. No. 1 le-09; Matches 110; Conservative 50; Mismatches 113; Indels 131; Gaps 14;

Qy 87 VCVQVAVAVRVAIAETLGLAGATTPQKLAEGRLRGGDEVDTLFGADLTIMEEGBLHLTEHIEAHPHS 161

Db 110 TPEKOLABGLR-----RLRDEVDTLFGADLTIMEEGBLHLTEHIEAHPHS 161

Qy 114 -----LADLDRFCONEMKTFKNAKYVYGTQFGKRSISISRINATLDRVPEN----- 159

Db 162 PLBMETSCPGWIAALERSYPDLYVSSCKSPQMLAMVLSYLAKKGIRKDWTAVS 221

Qy 160 EGPILCSVCPGEVLYAETKTPPELPHMLDVKSPOQITGNLHQ-----ADPTCYHLS 211

Db 222 IMPCTRQSEADBRWFCVDADPTLRLQDHVITVGLGNIFKERNIL-----AEL 271

Qy 212 IMPCFDKKLEASR-----ECEKEKDVYTRPKQFVAMGLDISDKSYMEYDSKEL 264

Db 272 PEGWDNFMVGSGAGVLFGTGTVGMEALAR-----TAYELFTG 310

Qy 265 CPGWDYKHLWSNEG--SSSGGYATOILSLSQSSNPESDITIEGKNSDTEYRLVSK 321

Db 311 TPFLRLSEVGRGMDGI-----KETNTIMWPAGPSKSEELLKHKRAAERAEEAHGTGCP 365

Qy 322 SKGVIASSESVYGRNTHONLVRKLSQSASVKKRGIK-----VKKRGQSVLK----- 367

Db 366 AMDGAGGTSEDEGGITLVRVAVANGLNNAKKLTKMGEAKYDFEIMACPGACVGG 425

Qy 368 -----SGTSE-----KTKVLT-----ADPKTDFEVMACPGCING 402

Db 403 GL-LNEEKNANRKKQLA-QDLSLAIYTKVHSVNIPDIVHYD 441

RESULT 15

HODVFS

hydrogenase (EC 1.18.99.1) (Fe) small chain precursor - Desulfovibrio vulgaris (strain

N) Alternative names: Fe hydrogenase beta chain

C;Species: Desulfovibrio vulgaris

C;Accession: B24551; A27480

R;Vooroud, G.; Brenner, S.

Eur. J. Biochem. 148, 515-520, 1985

A;Title: Nucleotide sequence of the gene encoding the hydrogenase from Desulfovibrio

A;Reference number: A24551; MUID:85203856; PMID:3888621

A;Accession: B24551

A;Molecule type: DNA

A;Residues: 1-123 <VOO>

A;Cross-references: GI:20416; NID:940829; PIDN:CAA26267.1; PID:940831

R;Prickett, B.C.; He, S.H.; Li, C.; Menon, N.; Choi, E.S.; Przybyla, A.E.; DerVartanyan, R.

Eur. J. Biochem. 149, 369-377, 1987

A;Title: Identification of three classes of hydrogenase in the genus, Desulfovibrio.

A;Reference number: A27480; MUID:88106446; PMID:332275

A;Accession: A27480

A;Status: preliminary

A;Molecule type: protein

A;Residues: 35-69 <PR1>

C;Comment: Three distinctive types of hydrogenases, the Fe, NiFe, and NiFeSe, are found

C;Genetics:

C;Complex: heterodimer of large (alpha, see PIR:HQDVF1) and small (beta) chains

C;Function:

A;Description: catalyze reactions involving the production or consumption of molecule

A;Note: may be involved in hydrogen uptake for the reduction of sulfate to hydrogen s

C;Superfamily: hydrogenase (Fe) small chain

C;Keywords: hydrogen metabolism; iron-sulfur protein; oxidoreductase; periplasmic spa

F;1-34/Domain: signal sequence #status predicted <SIG>

F;35-123/Product: hydrogenase (Fe) small chain #status predicted <MAT>

Query Match 9.0%; Score 232; DB 2; Length 469;

Best, Local Similarity	38.7%	Pred.	No.	0.0068	Matches
QY	429	RSTDKATIQKQRQ	-----	-----	29;
Db	27	RMTGKAVAVKQIKYMLDRINGVYGA	-----	-----	Conservative
QY	480	LLHITYV - AGGYEE	492	-----	Mismatches
Db	87	LLHITWFDKSKYKE	101	-----	-----

Search completed: June 3, 2003, 16:17:10
Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

On protein - protein search, using sw model
Run on: June 3, 2003, 16:07:42 ; Search time 22 Seconds

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Title: CAC80065
Perfect score: 2576
Sequence: 1 MSALVLPKPCAAVSTRGSSCR..... HELLHTHYVAGGVVEKDERK 497

Scoring table: BLOSUM62
Gapext 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	743.5	28.9	574	1 PHFL_CLOPA
2	633	24.6	421	1 PHFL_DESPTH
3	595.5	23.3	421	1 PHFL_DESTO
4	255	9.9	491	1 NARI_YEAST
5	232	9.0	469	1 LETL_KLUA
6	138.5	5.4	123	1 PHFS_DESPTH
7	130	5.0	124	1 PHFS_DESTO
8	112	4.3	776	1 HYPF_AZOT
9	111	4.3	912	1 TCA5_RABIT
10	109.5	4.3	857	1 CLP815_ECOLI
11	108.5	4.2	1887	1 FAS2_YEAST
12	107.5	4.2	391	1 ALR_STROCO
13	106	4.1	1885	1 FAS2_CANAL
14	105.5	4.1	3491	1 ERYL_SACER
15	105	4.1	447	1 ODP2_RHIME
16	105	4.1	896	1 IF2_CHLML
17	104	4.0	391	1 THIL_PARDE
18	103.5	4.0	377	1 PROB_METRU
19	103.5	4.0	1799	1 LMB2_MOUSE
20	102	4.0	835	1 RIRL_PRVKA
21	100	3.9	1842	1 FAS2_SCHPO
22	99	3.8	479	1 DPO4_STRCO
23	99	3.8	1322	1 XYLFLA
24	98.5	3.8	395	1 PURK_SYNPT
25	98.5	3.8	871	1 AR56_NEUCR
26	98.5	3.8	3695	1 LMA5_HUMAN
27	97	3.8	388	1 ALR2_AGR05
28	97	3.8	454	1 UCR4_NEUCR
29	96.5	3.7	732	1 YF48_HUMAN
30	96	3.7	1181	1 PIB2_HUMAN
31	95.5	3.7	553	1 SPAL_RAT
32	95.5	3.7	686	1 EFG2_STRCO
33	95.5	3.7	810	1 SYFP_SYNPT

ALIGNMENTS

RESULT 1
PHFL_CLOPA
ID PHFL_CLOPA
STANDARD: PRM: 574 AA.
AC P29166;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1999 (Rel. 28, Last annotation update)
DE Periplasmic [Fe] hydrogenase 1 (EC 1.18.99.1).
OS Clostridium pasteurianum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1501;
RN {1}
RN RP
RN RP
RN X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RN MEDLINE=9055388; PubMed=9836629;
RN Peters J.W., Lanzilotta W.N., Lemon B.J., Seefeldt L.C.;
RN "X-ray crystal structure of the Fe-only hydrogenase (CpI) from
RN Clostridium pasteurianum to 1.8 Å resolution.";
RN Science 282:1853-1858(1998).
RN -1- CATALYTIC ACTIVITY: 2 reduced ferredoxin + 2 H(+) = 2 oxidized
RN ferredoxin + H(2).
RN -1- COFACTOR: THIS PROTEIN BINDS PROBABLY FOUR 4FE-4S CLUSTERS AND ONE
RN H CLUSTER (POSSIBLE SIX-TO-EIGHT IRON CLUSTER).
RN -1- SUBUNIT: MONOMER.
RN -1- SUBCELLULAR LOCATION: Periplasmic.
RN -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
RN 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
RN -1- SIMILARITY: TO THE D. VULGARIS (HILDEBOROUGH) HD-GAMMA GENE
RN PRODUCT.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
--
EMBL: MB1737; AAA23248; 1;
DR PIR: A40330; HOCLIP.
DR PDB: 1FER; 06-JAN-99.
DR InterPro: IPR01150; 4Fe4S ferredoxin.
DR InterPro: IPR01150; 4Fe4S ferredoxin.
DR InterPro: IPR003149; Fe_hyd_ssU.
DR InterPro: IPR04108; Fe_hyd_1g_C.
DR InterPro: IPR0101; Ferredoxin.
DR Pfam: PF00037; fer4; 2.
DR Pfam: PF00111; fer2; 1.
DR Pfam: PF02256; Fe_hyd_SSU; 1.
DR Pfam: PF02905; Fe_hyd_1g_C; 1.

DE Periplasmic [Fe] hydrogenase small subunit precursor (EC 1.10.99.1)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-MAY-1991 (Rel. 18, Last annotation update)
 DE Periplasmic [Fe] hydrogenase small subunit precursor (EC 1.10.99.1)
 DE (Fe hydrogenlyase small chain).
 GN HYDB.
 OS Desulfovibrio vulgaris (strain Hildenborough).
 OC Desulfovibrio; Proteobacteria; delta subdivision; Desulfovibrionaceae;
 OC Desulfovibrio.
 OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
 OC Desulfovibrio.
 OX NCBI_TaxID:882;
 RN [1]
 RP SEQUENCE FROM N. A.
 RX MEDLINE=82203856; PubMed=3888621;
 RA Voordouw G., Brenner S.;
 RT "Nucleotide sequence of the gene encoding the hydrogenase from
 Desulfovibrio vulgaris (Hildenborough).";
 RL Eur. J. Biochem. 148:515-520(1985).
 RN [2]
 RP SIGNAL SEQUENCE CLEAVAGE SITE.
 RX MEDLINE=86277938; PubMed=3525521;
 RA Prickril B.C., Czechowski M.H., Przybyla A.E., Peck H.D. Jr.,
 RA Le Gall J.;
 RT "Putative signal peptide on the small subunit of the periplasmic
 hydrogenase from Desulfovibrio vulgaris.";
 RL Bacteriol. 167:722-725(1986).
 CC -I- FUNCTION: MAY BE INVOLVED IN HYDROGEN UPTAKE FOR THE REDUCTION OF
 CC SULFATE TO HYDROGEN SULFIDE IN AN ELECTRON TRANSPORT CHAIN.
 CC CYTOCHROME C3 IS LIKELY TO BE THE PHYSIOLOGICAL ELECTRON CARRIER
 CC FOR THE ENZYME.
 CC -I- CATALYTIC ACTIVITY: 2 reduced ferredoxin + 2 H(+) - 2 oxidized
 CC ferredoxin + H(2).
 CC -I- SUBUNIT: HETEROODIMER OF A LARGE AND A SMALL SUBUNIT.
 CC SUBCELLULAR LOCATION: Periplasmic.
 CC -I- MISCELLANEOUS: [FE], [NIFE], AND [NIFSE] HYDROGENASES APPEAR TO
 CC REPRESENT THREE DISTINCT ENZYMES HAVING HYDROGENASE ACTIVITY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X02416; CAA26267.1;
 DR PIR; B24551; HODVES.
 DR Interpro; IPR003149; Fe_hyd_SSU.
 DR Pfam; PF02256; Fe_hyd_SSU; 1.
 KW Oxidoreductase; Periplasmic; Iron-sulfur; Signal.
 FT SIGNAL 1 34 PERIPLASMIC [FE] HYDROGENASE SMALL
 FT CHAIN 35 123 SUBUNIT.
 FT SEQUENCE 123 AA; 13624 MW; 2F4F7A4304ECD47B CRC64;
 SQ
 Query Match 5.4%; Score 13.85; DB 1; Length 123;
 Matches 29; Conservative 14; Mismatches 21; Indels 11; Gaps 2;
 QY 429 RSTDKAETOKRQA-----ALYNDEKSTLRSHENPSPRELDTYLGEPGLKHA 479
 DB 27 RMTGKAVAVQKQIDMDRNGVYGADAKFPPVRSQDNTOVKALKSYLERPLGRKSHD 86
 QY 480 LLHHTHVV-AGGVEE 492
 DB 87 LLHHTHWFDSKSGVKE 101
 RESULT 7
 PHFS_DESVO STANDARD; PRT; 124 AA.
 PHFS_DESVO STANDARD; PRT; 124 AA.
 AC P13628;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAY-1991 (Rel. 18, Last annotation update)
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=DW;
 RX MEDLINE=92305060; PubMed=1610901;

RA Chen J C.; Mortenson L E.;
 RT "Identification of six open reading frames from a region of the
 RT *Acetobacter* vineelandii genome likely involved in dihydrogen
 metabolism";
 RL Blochim. Biophys. Acta 1131:199-202(1992).
 CC -I- FUNCTION: INVOLVED IN THE HYDROGENASE MATURATION PROCESS (BY
 CC SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE HYPF FAMILY.
 CC -I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
 ACYLPHOSPHATASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC

DR EMBL; X63650; CAA45185_1; ALT_INIT.
 DR L23970; AAA9510_1; ALT_INIT.
 DR InterPro; IPR001792; Acylphosphatase.
 DR InterPro; IPR004421; HyPF.
 DR InterPro; IPR00666; Suas_Yc10_yrdC.
 DR Pfam; PF00708; Acylphosphatase; 1.
 DR Pfam; PF01300; Suas_Yc10_yrdC; 1.
 DR Prodom; P0001884; Acylphosphatase; 1.
 DR Prodom; P002209; Suas_Yc10_yrdC; 1.
 DR TIGRFAMS; TIGR00143; hyPF; 1.
 DR PROSITE; PS00150; ACYLPHOSPHATASE_1; 1.
 DR ZINC-finger.

FT DOMAIN 1 93 ACRYLPHOSPHATASE-LIKE.
 FT ZN_FING 113 138 C4-TYPE (POTENTIAL).
 FT ZN_FING 163 188 C4-TYPE (POTENTIAL).
 SQ SEQUENCE 776 AA: 82732 MW: E99337COA15600C CRC64;

Query Match 4.3%; Score 112; DB 1; Length 776;
 Best local similarity 21.8%; Pred No 2.5%;
 Matches 130; Conservative 53; Mismatches 220; Indels 192; Gaps 28;

Oy 29 PLAATSVRVALATEAPARLIGN--VACAAAPAAEAPIS-----HV-----QQLA 73
 Db 182 PIACPACGPRCCEDASGRRLSGDPIERALALARGEITALRGVGGFHLACDARNAGAWA 241
 Oy 74 ELAMPKDDTRKHCVQVAPAVRAIAE-TLGLA- PGATTPQKLAEGRLRGDFEVD 129
 Db 242 ELRRRKRRRAKPFALMAANPASQALVENVNGLAELSPAAPVYLRL--KRPADDIL 299
 Oy 130 TLFGADLTMEEGEBLLRULTEHIEAPHSDEPLAUMTSCCPGVITAMLERKSYPDLLPV 189
 Db 300 AGVARDLAWL--GAMLPHSPHLWHLFHEAGR-----AGTGMA----- 337
 Oy 190 SCKSPQOMMAAMVSYLASKKGLAPKDMVNSIMPCTRKPOSEADR-----DWFCVADP 243
 Db 338 --APDILLIVMTSANLG-----AP-----PITGKAEARERLAGTADLWILHDKR 381
 Oy 244 TLRQDHWITVLEGNIFFERGINIAELPEGEWNNDPMGMSGAGVLEGTTGGIMEAELRT 303
 Db 382 LNRCDDSVSVALGRAPL-----EPLAIG-GPSIL--ALGG---QLRN 429
 Oy 304 AYELFTGTPPLPRLSEVIRG-----MDGKETNTMVFAP----- 338
 Db 430 AVCLTRG--DRAWLSPHNGDLDADTCALERVARLGELLGIRPERVACDLPDFPAS 486
 Oy 339 -----GSKEFELLKHARAARAAEAHAHGTGPG--LAWDG-GAERTSEDG--RGGI 382
 Db 487 RFARDYAERHGGLPLPQIHHHAATMAEHGLAEPVIGLALDQGFGLG--ADGRLRGE 543
 Oy 383 TLRVA-----VANGLNNAKKLTKMQGEAKDF 411
 Db 544 LLRVVAADGCAWLGLGELNPLPLPGGIDQASRPFWRMAGALHALGRGEIATRFAAEPGNAAI 603
 Oy 412 VEIMA---CP-----AGCIVGGGGOPRSTDRAITO-----KROALIN 446

RESULT 9

ICAS5_RABIT	STANDARD	PRT	912 AA.
ID ICR5_RABIT			
AC Q8730;			
DT 16-OCT-2001 (Rel. 40, Created)			
DT 16-OCT-2001 (Rel. 40, Last sequence update)			
DT 16-OCT-2001 (Rel. 40, Last annotation update)			
DE intercellular adhesion molecule-5 precursor (ICAM-5) (telencephalin).			
GN ICAM5 OR TICN.			
OS Oryctolagus cuniculus (Rabbit).			
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagidae; Oryctolagus.			
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagidae; Oryctolagus; NEIL TAXID=9986;			
RN [1].			
RP SEQUENCE FROM N.A.			
RC STRAIN-Japanese white; TISSUE=Brain;			
RX MEDLINE:94206534; PubMed=7794412;			
RA Yoshihara Y., Oka S., Nemoto Y., Watanabe Y., Nagata S.,			
RA Kogamiyama H., Mori K.,			
RT "An ICAM-related neuronal glycoprotein, telencephalin, with brain			
RT segment-specific expression";			
RL Neuron 12:541-553(1994).			
CC -I- FUNCTION: ADHESION MOLECULE THAT BINDS TO LEUKOCYTE ADHESION			
CC -I- LFA-1 PROTEIN (INTEGRIN ALPHAI-1/BETA1-2).			
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.			
CC -I- TISSUE SPECIFICITY: Expressed on neurons in the most rostral			
CC segment of the mammalian brain, the telencephalon, with brain			
CC -I- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. ICAM			
CC SUBFAMILY.			
CC -I- SIMILARITY: CONTAINS 9 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration - between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
DR EMBL; L13199; AAA18478_1;			
DR RSS; PF03362; IICL.			
DR InterPro; IPR03988; ICAM.			
DR InterPro; IPR03987; ICAM_VCAM-1.			
DR InterPro; IPR03006; Ig_MHC.			
DR InterPro; IPR03598; Ig_C2.			
DR InterPro; IPR03600; Ig-like.			
DR PRINTS; PRO1473; ICAM.			
DR PRINTS; PRO1472; ICAMVCAM1.			
DR SMART; SM00410; Ig_Like_3.			
DR SMART; SM00408; IgC2; 1.			
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Transmembrane;			
KW Repeat; Signal.			
FT SIGNAL 1 29 POTENTIAL.			
FT CHAIN 30 912 INTERCELLULAR ADHESION MOLECULE-5.			
FT DOMAIN 30 826 EXTRACELLULAR (POTENTIAL).			
FT TRANSMEM 827 847 POTENTIAL.			
FT DOMAIN 848 912 CYTOPLASMIC (POTENTIAL).			
FT DOMAIN 849 107 Ig-LIKE C2-TYPE DOMAIN 1.			
FT DOMAIN 132 202 Ig-LIKE C2-TYPE DOMAIN 2.			
FT DOMAIN 239 304 Ig-LIKE C2-TYPE DOMAIN 3.			
FT DOMAIN 323 385 Ig-LIKE C2-TYPE DOMAIN 4.			
FT DOMAIN 403 472 Ig-LIKE C2-TYPE DOMAIN 5.			
FT DOMAIN 486 552 Ig-LIKE C2-TYPE DOMAIN 6.			
FT DOMAIN 566 645 Ig-LIKE C2-TYPE DOMAIN 7.			

RESULT 10
CLPB_ECOLI
ID CLPB_ECOLI STANDARD;

PRT; 857 AA-

R1

"Expression of ClpB, an analog of the ATP-dep-

10

RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M., Chung E., Churcher C.M., Coster F., Davis K., Davis R.W., Dietrich F.S., Delius H., DiPolo T., Dubois E., Duysterhoeft A., Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A., Hahn J., Hebling U., Heumann K., Hilbert H., Hillier L., Hunnicut-Smith S., Hyman R., Johnston M., Kalman S., Kleine K., Komp C., Kundi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J., Marathie R., Messinguy F., Mewes H.-W., Mirtipati S., Moestl D., Mueller-Auer S., Namath A., Neutzeberg U., Oefner P., Pearson D., Petel F.X., Pohl T.M., Purnelle D., Schaefer M., Scharfe M., Scheipers B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H., Urrestarazu A., Ushinsky S., Vierendeels F., Vissers S., Voss H., Walsh S.V., Wambutt R., Wang Y., Wedler E., Weidner H., Winnett E., Zhong W.W., Zolnier A., Vo D.H., Hani J.;
RA "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI.";
RA *Nature* 387:103-105(1997).

RA [4] MUTAGENESIS OF GLY-1250.

RA RN .

RA RP .

RA RC .

RA STRAIN=S288C;

RA RX MEDLINE=9416198; PubMed=8041367;

RA RA Inokoshi J., Tomoda H., Hashimoto H., Watanabe A., Takeshima H., Omura S.;
RA "Cerulenin-resistant mutants of *Saccharomyces cerevisiae* with an altered fatty acid synthase gene.";
RA RT Mol. Gen. Genet. 244:90-96(1994).

RA RL .

RA CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH. THE ALPHA SUBUNIT CONTAINS DOMAINS FOR: ACYL CARRIER PROTEIN, 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND 3-OXOACYL-[ACYL-CARRIER PROTEIN] SYNTHASE. THIS SUBUNIT COORDINATES THE BINDING OF THE SIX BETA SUBUNITS TO THE ENZYME COMPLEX.

RA CC -1- CATALYTIC ACTIVITY: ACETYL-COA + N malonyl-COA + 2N NADPH = a long-chain fatty acid + (N+1) COA + N CO(2) + 2N NADP(+).

RA CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) + [acyl-carrier protein].

RA CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] + NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADH.

RA CC -1- SUBUNIT: [Alpha(6)beta(6)] hexamers of two multifunctional subunits (alpha and beta).

RA CC -1- SIMILARITY: TO THE FATTY ACID SYNTHETASE, SUBUNIT ALPHA FROM OTHER FUNGI.

RA CC .

RA This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

RA CC .

RA DR J03936; AAA34601.1; .

RA DR EMBL; J03936; AAA34601.1; .

RA DR EMBL; X76890; CA054218.1; .

RA DR EMBL; X94561; CA064256.1; .

RA DR EMBL; 273586; CA0947.1; .

RA DR EMBL; 273587; CA097948.1; .

RA DR PIR; A31107; A31107.

RA DR SEDB; S0006152; FRS2.

RA DR InterPro; IPR002582; ACPS.

RA DR InterPro; IPR000794; Ketoacyl-synt.

RA DR InterPro; IPR004568; Pantethin_trn.

RA DR InterPro; IPR030380; Ppantne_attach.

RA DR pfam0109; ketoacyl-synt; 1.

RA DR pfam; PF01648; ACPS; 1.

RA DR PDB02801; ketoacyl-synt_C; 1.

RA DR Prodrom; PD004282; ACPS; 1.

RA DR TIGRFAMs; TIGR00556; pantethin_trn; 1.

RA DR PROSITE; PS00012; PHOSPHOPANTETHINE; 1.

RA DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.

RA KW fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase; KW transferase; NADP; Phosphopantethine.

RA DOMAIN 1 ACYL CARRIER (ACP);
RA DOMAIN 2 ACYL-KETOACYL REDUCTASE.

RA DOMAIN 675 874

RA [5] "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2);"
RA "Nature" 417:141-147(2002).

RA -1- FUNCTION: Provides the D-alanine required for cell wall biosynthesis (By similarity).

RA CC -1- CATALYTIC ACTIVITY: L-alanine = D-alanine.

RESULT 15

OP2-RHIME STANDARD; PRT; 447 AA.

ID ODP2-RHIME

AC 09R9N3;

DT 30-MAY-2000 (Rel. 39, Created)
15-JUN-2002 (Rel. 41, Last sequence update)

DE Dihydropalpamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12) (E2).

DE PDB; OR PDB; OR RSCM0132;

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiales group; Rhizobiaceae; Sinorhizobium.

OC NCBI_TaxID=382;

OX RN

SEQUENCE FROM N.A.

STRAIN=RCR2011 / SU47;

RX MEDLINE=20355031; PubMed=10796014;

RX Cabanes D., Boistard P., Batut J.,

RX "Symbiotic induction of pyruvate dehydrogenase genes from Sinorhizobium meliloti."

RT Mol. Plant Microbe Interact. 13:483-493 (2000).

[2]

RR SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;

RX Capelard D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J., Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S., Godrie T., Goffeau A., Kahn D., Kiss E., Lelauze V., Masy D., Pohl T., Portetelle D., Puchier A., Purnelle B., Ransperger U., Reznor C., Thebaud P., Vandebil M., Weidner S., Galibert F.; "Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021." J. Proc. Natl. Acad. Sci. U.S.A. 98:9177-9822 (2001).

RT -1- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1), Dihydropalpamide acetyltransferase (E2) & LIPOAMIDE DEHYDROGENASE (E3) (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: Acetyl-COA + dihydropalpamide = COA + S- acetyl-dihydropalpamide.

CC -1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL (BY SIMILARITY).

CC -1- SUBUNIT: FORMS A 24-POLEPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL SYMMETRY (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.

CC -1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; AF190792; AAF05899; 1; -;

DR EMBL; AL591787; CAC46026; 1; -;

DR HSSP; P07016; IBBL.

DR InterPro; IPR01078; 2oxoacid_dh.

DR InterPro; IPR00089; Biotin_lipoyl.

DR InterPro; IPR04167; E3_binding.

DR InterPro; IPR03016; Lipoyl.

DR Pfam; PF00198; 2-oxoacid_dh; 1.

DR Pfam; PF00364; biotin_lipoyl; 1.

Query Match 41%; Score 105; DB 1; Length 447;

Best Local Similarity 19.5%; Pred. No. 39; Mismatches 161; Indels 190; Gaps 23; Matches 99; Conservative 57; Mismatches 161; Indels 190; Gaps 23; Matches 99;

Qy 11 AVSIKSSCRARQVAPRPLAASVRLAATLEPARRRGINVACAAAPAPAPL-----

Db 85 ATAAKGNGAAGAVP--APKKETAAAPAAPA--AAPAPAPAAPAAPAPASPADEGK 139

Qy 66 -----SHVQQALAE-----LAKPKDDDTTRKHCVQVAPAVRVAIAETIGLPGATTPQL 115

Db 140 RIFSSLPLARRLAKEAGIDLSAISAGSGHGRVVKDQETAVSGGAAKPAGPAPAPAPL 199

Qy 116 AEGLRRLGDFDEVDTLFGADLTIMEGCS-----ELLRLTELEAHPSDDEPLP 164

Db 200 AKGME-----DAVKLIFEPGSYLYPHDGMRTKIAKRLVSKQTIHPF---- 243

Qy 165 MFTSCCPGWIAMEKSYPDLTIVYS-SCKSPGDM-LAMVKSIAEKG-----IAPKD 217

Db 244 -----YVSVDCDCELDAMALRAQNLAAAPEKDKGPKVYLSDM 281

Qy 218 WMSVIMPCTRKSEADRWFCVDADPILRQLDHWITVVELGNIFKERGINLALBPEGWD 277

Db 282 VIKALALALRQVDPANWS----TDQNNVKKH----- 310

Qy 278 NPMGVGSGAGVLTGFTGGMVMEALRTAYELFTGCTPLPRLSLEVRGMDGKIKETNITWPA 337

Db 311 -----ADVGAVAVSIPGGLI-----TPI-----VROAELKLSA 338

Qy 338 PCKSFELLIKHAAAREAAAHGTPGFLAWGGAGFTSEDRGGITLRAVAV-GLGNAK 396

Db 339 ISNEMKDLGK-RAKERK-----LKEPYOGTT---AVSNGMMGVK 376

Qy 397 KLTIMQAGEAKYDFEVETMACPAG--CVGGQQPRSTDQKATQKQRAQLYNIDESTL 453

Db 377 -----DFAAVNPPHATLAVGAG---EDRVVWRKENVANY-MTVLS 417

Qy 454 RSHE-----NPSIRELYDTYLGEPPLG 474

Db 418 TDHRCVDGALGAGELLAFAKRYIENPMG 444

Search completed: June 3, 2003, 16:16:21
Job time : 25 secs